DRDC | RDDC
technologysciencetechnologie

Challenges of sentinel testing for early detection of coronavirus disease 2019 (COVID-19) infections

Application to office buildings or other similar environments

Ramzi Mirshak
DRDC – Centre for Operational Research and Analysis

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2020-R127
December 2020

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by DRDC using the Schedule to the *Defence Production Act*.

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

This Scientific Report examines the utility of regular sentinel testing for monitoring coronavirus disease 2019 (COVID-19) activity in an office building or other similar setting where a group of individuals congregate but also mix with the general population. This is different from traditional sentinel testing where the objective is to understand the prevalence of a large region. Here, the goal is to ascertain the health of the building population relative to that regional scale.

Two methods are considered. The first is known as Freedom from Disease (FFD), which was developed in veterinary medicine to examine livestock infections across a large region and was recently suggested as a tool to demonstrate absence of COVID-19 on a military base. The second is an extension that accepts the virus may be within the group and, considering the effect of false positives, provides insights on the point prevalence in the group relative to the background point prevalence of the community. Both methods apply a Bayesian approach to make these inferences.

Due to the mixing of the group of interest with the general population, the probability of introduction of the virus between rounds of testing is quite high. As a result, unless testing is at extremely high levels, it is not possible to get a high confidence that the point prevalence in the group is low. As a result, neither method suggests sentinel testing will provide much protection for or information about the group.

Additional investigation shows that the time-dependent aspect of testing, i.e., that an individual must be tested in a timely manner for sentinel testing to be effective, further exacerbates the problem. Ultimately, true positive tests from sentinel testing are likely to be overwhelmed both by false positives and by individuals who are identified clinically after symptoms present.

Although exploring other options is beyond the scope if this work, mitigation measures that limit spreading opportunities before an infection is known and aggressive contact tracing after an infection is identified are likely to be more effective at protecting the group from an outbreak than a limited sentinel sampling strategy.

Significance for defence and security

Due to the need to work with classified or other guarded information or material, the requirement for individuals working in military and defence establishments to work in close proximity may be high. Whether this be on a base or in a headquarters building, many will live at home and mix with the general population. Sentinel sampling has been and continues to be used at some establishments. Understanding the limitations of a sentinel testing strategy can inform decisions on how to best mitigate against COVID-19 in these environments, thereby applying limited resources in a manner that best protects the Force and continuity of efforts.

Résumé

Ce rapport scientifique examine l'utilité des tests sentinelles réguliers pour surveiller l'activité de la maladie à coronavirus 2019 (COVID-19) dans un immeuble de bureaux ou autre lieu similaire où les personnes se regroupent, mais se mélangent aussi avec la population générale. Cela diffère des tests sentinelles traditionnels dont l'objectif est de comprendre la prévalence de la maladie dans une grande région. Dans le présent document, l'objectif est de vérifier la santé de la population de l'immeuble par rapport à celle de la population à l'échelle régionale.

Deux méthodes sont envisagées. La première méthode est connue sous le nom de validation de l'absence de la maladie, qui a été élaborée en médecine vétérinaire pour examiner les infections du bétail dans une grande région et a récemment été proposée comme outil pour démontrer l'absence de COVID-19 dans une base militaire. La deuxième méthode est une extension qui accepte que le virus puisse se trouver dans le groupe et, compte tenu de l'effet des faux résultats positifs, donne des indications sur la prévalence instantanée dans le groupe par rapport à la prévalence instantanée de base de la communauté. Les deux méthodes appliquent une approche bayésienne pour faire ces déductions.

En raison du mélange du groupe d'intérêt avec la population générale, la probabilité d'introduction du virus entre les séries de tests est assez élevée. Par conséquent, à moins que le nombre de tests effectués soit extrêmement élevé, il n'est pas possible d'obtenir une grande certitude que la prévalence instantanée dans le groupe est faible. Par conséquent, aucune des deux méthodes ne suggère que les tests sentinelles apporteront une grande protection du groupe ou des renseignements sur le groupe.

Une enquête supplémentaire montre que l'aspect temporel des tests, c'est-à-dire qu'une personne doit être testée en temps utile pour que les tests sentinelles soient efficaces, aggrave encore le problème. En définitive, les vrais résultats positifs des tests sentinelles risquent d'être dépassés à la fois par les faux résultats positifs et par les personnes qui sont testées cliniquement après la présence des symptômes.

Bien que l'exploration d'autres options dépasse le cadre de ces travaux, les mesures d'atténuation, qui limitent les possibilités de propagation avant qu'une infection ne soit connue et la recherche agressive des contacts après l'identification d'une infection, seront probablement plus efficaces pour protéger le groupe contre une éclosion qu'une stratégie d'échantillonnage sentinelle limité.

Importance pour la défense et la sécurité

Étant donné qu'il leur faut travailler avec des informations ou du matériel classifiés ou protégés, l'obligation pour les personnes travaillant dans les établissements militaires et de défense de travailler à proximité immédiate peut être élevée. Que ce soit dans une base ou dans un bâtiment du quartier général, beaucoup d'entre eux vivront chez eux et se mêleront à la population générale. L'échantillonnage sentinelle a été et continue d'être utilisé dans certains établissements. La compréhension des limites d'une stratégie d'échantillonnage sentinelle peut éclairer les décisions sur la meilleure façon d'atténuer les effets de la COVID-19 dans ces environnements, ce qui permet d'utiliser des ressources limitées de manière à protéger le mieux la Force et la continuité des efforts.

Table of contents

List of figures

List of tables

Acknowledgements

The author would like to thank Jennifer Steele of United States European Command for early discussions about the sampling strategies used in veterinary medicine. Her understanding of Freedom from Disease and its application was extremely helpful.

Steve Schofield of Canadian Forces Health Services participated in deep discussions about test sensitivity, which helped shape the section considering the time-dependent aspect of the problem.

Discussions about this and related work with François Cazzolato, Peter Dobias, Steven Horn, Andrew Sirjoosingh, David Waller, and others in DRDC – CORA have also been crucial for this effort. In particular, however, interactions with Steve Guillouzic on the methodologies related to point prevalence, incidence rate, and their relationship were instrumental in guiding the way forward in understanding the time-dependent aspects of the problem and properly defining the probability of new cases coming into the population.

Finally, the author thanks the anonymous peer reviewers who gave constructive comments that further improved this work, helping to frame its place relative to the concerns in the public health sphere.

1 Introduction

1.1 Background

With the coronavirus disease 2019 (COVID-19) epidemic enduring for almost a year so far and with it forecast to last for many more months at least, long-term isolation has not been feasible for all and it has been necessary for many individuals to return safely to work. While many are encouraged to work from home where possible, others are unable to do so. Some tasks simply require people to be present at their workplace, to be working collaboratively in relatively close proximity for example. In defence and military settings, the need to work with classified or other guarded information has limited the ability of many to work from home. This situation results in a tension between the need to bring people back into the workplace and the risk of creating the opportunity for a COVID-19 outbreak in that same workplace.

A number of behavioural changes in the workplace are known reduce the likelihood of an outbreak, including: reducing the number of people working at any time, breaking the work place into zones, restricting movement between zones, and limiting in-person meeting participation. While these interventions have the potential to reduce the likelihood and magnitude of an outbreak should an infected individual come into work, they are examined elsewhere [1, for example]. Similarly, the impact of mask wearing will further reduce the probability of spread [2]. Together with contact tracing, the above behaviours greatly reduce the viral transmission in the workplace [3, 4]. While the above interventions are all important, however, they are by no means an exhaustive list of mitigation measures and, in fact, remain beyond the scope of this report. Instead, this effort considers the degree to which a sentinel sampling strategy might help to ensure workplace population health.

Sentinel testing is an important surveillance tool for tracking disease prevalence and spread. As described by the World Health Organization (WHO), "[d]ata collected in a well-designed sentinel system can be used to signal trends, identify outbreaks and monitor the burden of disease in a community, providing a rapid, economical alternative to other surveillance methods." [5] However, this is not the entire story. Continuing, the WHO also includes a word of caution. Because the testing program only looks at select locations, "sentinel surveillance...may not be as effective for detecting rare diseases." This leaves an open question: when might a disease be considered rare, and where might the effectiveness of sentinel testing begin to weaken?

The focus of this effort is to examine what sentinel testing provides in terms of information on COVID-19 prevalence in a small group such as an office building, and whether it provides an ability to mitigate the spread of COVID-19 in that workplace. Because the problem at hand considers a relatively small population, the law of large numbers [6] that makes surveillance sampling informative starts to break down, and so the caveat of rare cases provided by the WHO [5] comes into play. In addition to exploring the efficacy of sentinel testing, given that one will need to contact trace and isolate individuals each time there is a positive test, the implications of false negatives are also explored. This is an important factor for the problem at hand since, assuming individuals are coming into work out of necessity, it affects how to best manage staff carrying out essential duties in the workplace to ensure redundancy. The analysis in this report is explored for a specific example scenario, where a workplace is comprised of individuals who are not isolated from but rather mix with their community. There is a constant background disease prevalence within this community, which in turn provides a constant risk of new infections for the group of interest.

1.2 Problem description and scope of effort

Figure 1 shows a depiction of the problem where a subset of individuals (beige stick figures) are part of a subpopulation that is working in close quarters, such as an office building (beige circle). As part of the broader community, these individuals are interacting with others and so remain at an assumed constant risk of contracting a severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) infection. While people in the general population are not likely to be infected (grey stick figures), a small fraction will be (red stick figures). Hence, an individual in the subpopulation of interest could become infected and unknowingly bring the virus into the workplace, thereby creating a potential to cause an outbreak there.

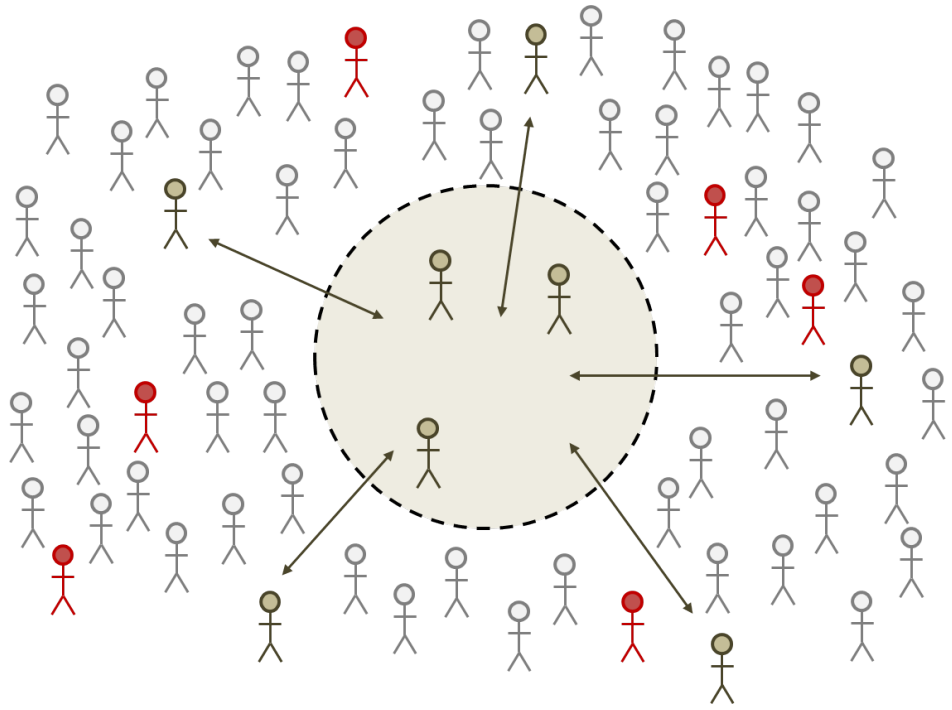


Figure 1: *Depiction of problem, where the goal is to use sentinel testing to ensure that the disease is controlled within the subpopulation of interest.*

The problem here is to understand what sentinel testing does and does not provide in terms of understanding the status of SARS-Cov-2 infections in the building at an instance in time, rather than to understand the number of people who have been infected over time. This is an important distinction as it affects how best to approach the problem. For example, if one wants to understand the fraction of the population that has become infected, then one could track the number of individuals who test positive for the virus over time. From there, one can make assumptions about the bias of testing towards a fraction of symptomatic individuals and thus infer the likely number of people who have been infected in total [7]. Alternatively, if enough people are testing positive within the population of interest, then one can infer the incidence rate and from that, the number of unknown infected individuals [3]. However, in scenarios where one is looking at a smaller group of people where few people are displaying symptoms over time, then these types of inferences may be less applicable. It is this situation in particular that is of interest in this report, and the question at hand is what sentinel testing brings to that problem under this circumstance. Sentinel testing is not designed to detect symptomatic individuals, as they should self-identify, but rather those who

are infected but not showing symptoms. Since one expects that there will be infections within the population of interest over time, another question to consider is whether sentinel testing can help to infer if there is spread within the population of interest.

Early in the pandemic, some proposed an approach to sentinel testing for military bases known as Freedom from Disease (FFD) [8], which is born out of veterinary epidemiology [9, 10]. The approach uses successive rounds of testing to demonstrate the probability that a group is free from disease. As a result, the probability of introduction of the disease to the group between rounds of testing is important. If tests give confidence that the point prevalence in the subpopulation of interest is low but there is also a high degree of mixing with the general population, then the value of those test results discounts quickly since the likelihood of people contracting the virus after testing increases quickly. If testing levels are too low or too infrequent, the mixing effect shown in Figure 1 will overwhelm test practices and testing will give little confidence A schematic showing the difference between having sufficient and insufficient testing in a sampling strategy is shown in Figure 2. In scenarios where testing is sufficient and the probability of the disease entering the group is low, the testing regime used in FFD aims to ensure a high confidence that the disease is not within the subpopulation of interest (Figure 2a). However, if testing is insufficient and the probability of the disease entering the group is high, then the testing regime will not provide much utility (Figure 2b).

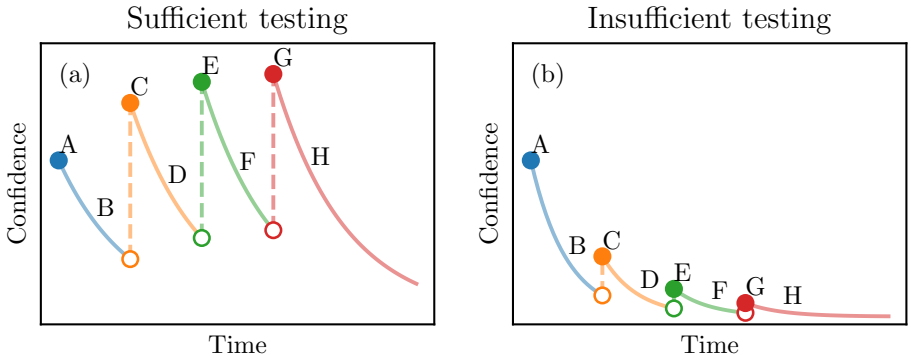


Figure 2: *Demonstration of confidence in sampling strategy given (a) sufficient testing; and (b) insufficient testing. Testing requirements are based on background point prevalence and group size. A—Initial belief/confidence; B/D/F/H—decrease in confidence due to mixing with population; C/E/G—increase in confidence due to testing.*

At the time that FFD was proposed [8] to monitor COVID-19 on a military base, the probability of introduction was not well constrained and it was not clear whether it would be sufficiently low for FFD to be a successful approach. Recently, other efforts [11, 12] have helped to estimate that aspect of the problem, and the methods developed are applied to incorporate a realistic estimate for the probability of introduction.

Another limitation of the FFD approach is that it assumes perfect specificity, i.e. no false positive results. The premise of FFD is that once testing reveals a positive test, the assertion that the group is free of disease is rejected. This causes two issues. First, it is unlikely that a group will remain disease-free *ad infinitum*, so at some point FFD can provide no information except that the disease is in the group. Second, a non-zero false positive rate is expected SARS-Cov-2, implying that the assumptions linked to FFD may not fit the problem at hand. To address these issues, this report

also develops an extension of the FFD approach where one accepts that the disease may be present in the population of interest. This extension considers the effect of false positives and attempts to estimate the prevalence in the group of interest relative to an estimate for the region. It will be shown that both FFD and the proposed extension are unlikely to provide much information concerning the status of COVID-19 infections in the workplace relative to the general population.

In spite of these shortcomings, some may still wish to push forward with a sampling strategy in order to at least catch some infections before they have an opportunity to spread within the workplace. Examining this question highlights another shortcoming of sentinel testing: not only does an infected individual need to be caught but they need to be caught *in a timely manner*. Symptomatic cases must be detected before symptoms present (otherwise the individual will presumably self-identify and hopefully get tested outside of the sentinel program), while asymptomatic cases must be detected while they are still infectious to minimize the chance of viral spread [13].[1] We find that not only will a sentinel strategy have difficulty detecting these infections, but also that even a mild false positive rate will dominate the signal. Assuming those in the workplace need to be there for essential purposes, and assuming that contact tracing and isolation of contacts will follow any positive tests, the implications of false positives also needs consideration. In particular, one would want to develop schedules that permitted continuity of essential functions while being prepared for multiple periods isolation.

1.3 Working example

Throughout this document, a generic example is used where testing is conducted for a building with $N = 500$ workers who live in a region with a point prevalence of $\pi_0 = 0.5\%$.[2] (Table 1 provides definitions for all symbols used in the body of the document.) In some cases this prevalence is prescribed as that exact value, while in others uncertainty is included and $\pi_0 = 0.5\% \pm 0.125\%$. In Sections 3 and 4, the test sensitivity is set to $Se = 0.7$, or 70% [14, 15], meaning the false negative rate is 30%. In Section 5, when considering a time-window during which sentinel testing is effective, perfect sensitivity (given sufficient viral load) is also considered to give an upper bound. For cases where specificity is imperfect, we assume a value of $Sp = 0.99$, meaning that 1% of negative cases test positive.[3] While test specificity may indeed be higher than 0.99 in reality, this value is used as a low bound to constrain the findings. Towards the end of the document in the discussion, some key results are reviewed for a wider range of specificity values, covering the range between 0.98 and 1.

To demonstrate the efficacy of sentinel testing, we employ a sampling strategy where 10% of the subpopulation, that is $k = 50$ individuals, are selected randomly for each week for testing ($\tau = 7$ days), although the impact of varying k and τ is considered in some scenarios.

[1] Asymptomatic cases detected after they are still infectious may still provide an ability to contain the spread of the disease through contact tracing. However, the degree to which this is likely to be effective is beyond the scope of the current report.

[2] This value is well below what has been witnessed El Paso County so far [7], but is well within the realm of realistic values across North America. Note that other regions in the United States continue to be above this value at the time of writing, and a second wave could well bring similar or higher values to urban centres in Canada.

[3] Sensitivity is defined as $Se = TP/(TP + FN)$ and specificity is defined as $Sp = TN/(TN + FP)$ where TN, TP, FN, and FP are the number of true negatives, true positives, false negatives and false positives, respectively.

Table 1: *Definitions of symbols and parameters used in main body of text.*

Symbol	Definition
d	Number of infections within a population
d^*	Expected number of infections within the population based on π_d
f_v	Cumulative density function of the probability of sufficient viral load for testing
h	Probability of being within the time window for testing at time t
k	Number of people tested
I	Indication that an infection is introduced to the group
N	Group size
$p(\cdot)$	Probability distribution of $(\cdot)$
p_t	Fraction of population tested in a test period, i.e. k/N
p_{test}	Probability that an infected individual is tested
p_{detect}	Probability that an infected individual is detected by the sampling strategy
p_a	Probability that an individual case will be asymptomatic
p_g	Probability of someone in a group becoming infected in the time period of interest
r	Incidence rate for new infections per person
r_g	Incidence rate for new infections in the group
Se	Test sensitivity
Se_s	Sensitivity of sentinel strategy
Sp	Test specificity
S_p	Survival function for the pre-symptomatic phase of symptomatic infections
S_i	Survival function for the infectious period of mild symptomatic cases
S_c	Survival function of all COVID-19 cases
S_{si}	Survival function of the symptomatic phase of symptomatic cases
S_a	Survival function of asymptomatic cases
t	Time
T^+	Number of positive tests
T_f^+	Number of false positive tests
y	Number of infected individuals selected for sentinel testing
α, β	Parameters describing the beta distribution
μ_o, σ_o	Centre and scale of f_v
μ_i	Mean duration of a SARS-Cov-2 infection
μ_g, σ_g	Parameters of the gamma function defining S_i
μ_p, σ_p	Parameters of the lognormal distribution defining S_p
μ_s, σ_s	Parameters of the gamma function describing S_{si}
π	Point prevalence
π_d	Design prevalence
$\hat{\pi}$	Inferred point prevalence in group
π_0	Inferred point prevalence in society
τ	Time period between tranches of sentinel testing

1.4 Document structure

This document is organized as follows. First, assumptions pertaining to virus pathology, test characteristics, and sentinel testing are outlined in Section 2. These assumptions underpin much of the work derived in the following three sections. Next, an outline of the concept of FFD and its limitations for the problem at hand are discussed in Section 3. As mentioned above, a limiting factor of this approach is that it requires all tests to come back negative, which implicitly assumes there are no false positives (i.e., perfect test specificity). To address this limitation, an alternative approach that aims to use sentinel testing to estimate the likelihood that the point prevalence in the population is at or below the background level is introduced in Section 4. This second approach also has limited utility and, while it can handle imperfect specificity, the ability to infer information from such a test is shown to be limited for the problem at hand.

Following presentation of these two methods, we examine the time-dependence aspect of the infection itself in Section 5 since, in order to be effective, the sampling strategy must catch infections before they develop symptoms (for symptomatic cases) or before they cease to be infectious (for asymptomatic cases). The report then finishes with a discussion in Section 6, which ties the results from Section 5 to those in Sections 3 and 4, explores implications of higher and lower specificity values, and outlines the author's assessment of the implications of the findings.

2 Assumptions

Estimating the daily probability of infection of individuals is an important parameter for estimating the risk of infection to individuals or groups over time and there have been several recent efforts to understand that problem, with each building on the previous [11, 12, 16]. In particular for the efforts here, important factors about the disease are those that relate to the time after infection before there is a sufficient viral load to test positive, the probability of different types of infections (i.e., symptomatic versus asymptomatic), and the duration of those infections. In addition, one must make assumptions with regards to the test characteristics, i.e., the sensitivity and specificity.

2.1 Test characteristics

1. When considering FFD in Section 3, test sensitivity is $Se = 0.7$ and test specificity is $Sp = 1$. For the prevalence modelled ($\pi = 0.5\%$), this is equivalent to a positive predictive value (PPV) of 1 and a negative predictive value (NPV) of 0.998.

2. When considering the probability of being at or below the background point prevalence in Section 4, test sensitivity is $Se = 0.7$ and test specificity is either $Sp = 0.99$ or $Sp = 1$. For the imperfect specificity, and for the prevalence modelled, this is equivalent to a PPV of 0.26 and a NPV of 0.998.

3. When considering the time-window for detecting an infection in Section 5, the ability to catch an infection early depends on the viral load (discussed further in Assumption 13). It is not entirely clear how this is factored into the sensitivity value used above, so bounds of $Se = 0.7$ and $Se = 1$ are both considered when including the window for testing. Specificity values of $Sp = 0.99$ and $Sp = 1$ are used and compared.

2.2 Sentinel sampling scenario

4. Sampling strategies are applied to an office building of $N = 500$ individuals, who are not quarantined or sequestered, and mix regularly with the general population (Figure 1).

5. Sampling strategies are assumed to select individuals at random, meaning the possibility of focused testing of specific potential high-risk individuals within the building is neglected.

6. Sentinel testing occurs weekly ($\tau = 7$ days), although the parameter space for different testing intervals is also explored.

7. In each round of sentinel testing, $k = 50$ individuals are selected randomly, although the parameter space for different proportions of the population is also explored. Due to the random nature of the strategy, testing someone in one round does not preclude them from being tested in any following rounds.

8. Within the group, there are presumed to be d infected individuals, which is to say that the point prevalence within the group is $\pi = d/N$.

In order to understand and interpret how a sampling strategy will play out, one must appreciate the probability of having x positive tests result from testing the k individuals. Because the population of interest N is relatively small, we apply a hypergeometric distribution rather than a binomial one to determine the probabilities.

9. For a random sampling strategy, the likelihood of drawing y infected individuals out of the population of N individuals follows:

$$p(y|N,k,d) = \frac{\binom{d}{y}\binom{N-d}{k-y}}{\binom{N}{k}}, \tag{1}$$

where $y \leq d$.

10. For the FFD scenario with perfect specificity and imperfect sensitivity, the likelihood of drawing x positive tests ($T^+ = x$) builds on Equation (1) [17]:

$$p(T^+ = x|N,k,d,Se,Sp = 1) = \sum_{y=x}^{d} \frac{\binom{d}{y}\binom{N-d}{k-y}}{\binom{N}{k}} \binom{y}{x} Se^x (1-Se)^{y-x}, \tag{2}$$

where $x \leq y \leq d$. The added terms on the right hand side of the equation give the binomial distribution that considers the likelihood of x out of y individuals testing positive.

11. For scenarios that also consider specificity, we add another component except this time to consider the possibility of a non-infected individual testing positive [17]:

$$p(T^+ = x|N,k,d,Se,Sp) = \sum_{y=0}^{d} \frac{\binom{d}{y}\binom{N-d}{k-y}}{\binom{N}{k}} \sum_{j=0}^{\min(x,y)} \binom{y}{j} Se^j (1-Se)^{y-j} \binom{k-y}{x-j} (1-Sp)^{x-j} Sp^{k-x-y+j}. \tag{3}$$

We apply Equation (1) when considering the probability of testing a given number of infected individuals and Equations (2) and (3) whenever considering probability of someone testing positive given the point prevalence, the size of the group, and the number of people being tested.

To assess the number of true and false positives relative to the number of actual cases in Section 5, it is necessary to identify the number of false positives and true positives. The number of true positives can be identified using Equation 2, but false positives requires a modification of Equation 3 where only false positives are considered.

12. For the number of false positives, we remove the true positives from the group sampled and explore the probability of getting false positives out of the remaining members of the group being tested:

$$p(T_f^+ = x|N,k,d,Sp) = \sum_{y=0}^{d} \frac{\binom{d}{y}\binom{N-d}{k-y}}{\binom{N}{k}} \binom{k-y}{x} (1-Sp)^x Sp^{k-x-y}, \tag{4}$$

where T_f^+ is the number of false positive test results.

2.3 Properties of SARS-Cov-2 infection

The key assumptions as they relate to the behaviour of COVID-19 infections are:

13. At the onset of infection, an insufficient viral load will lead to a negative test. It is only after the virus has had sufficient time to replicate in its host that one can expect a positive test. Results from [18] suggest that one is unlikely to test positive until several days after contracting the virus. For demonstration purposes here, and following the findings in Figure 3 from [18], we assume it takes two to three days after infection for the viral load to be sufficiently high to test positive. Modelling the probability of a sufficient viral load as a logistic function centred at $\mu_o = 2.5$ d and with a scale of $\sigma_o = 0.25$ d, the cumulative density function of the probability of having a sufficient viral load to test positive is,

$$f_v(t) = \frac{1}{1 + e^{-(t-\mu_o)/\sigma_o}}. \tag{5}$$

This assumption only comes into play in Section 5. To simplify the problem, we further assume that f_v is independent of the duration of the infection and of whether the case is symptomatic or asymptomatic.

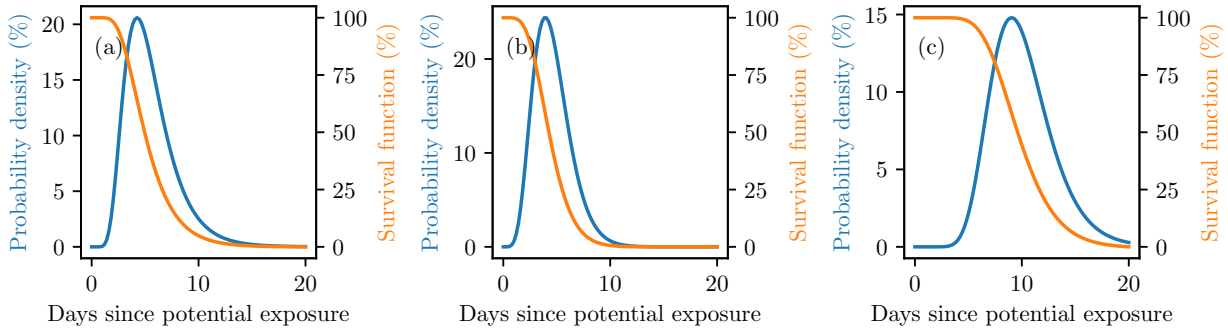


Figure 3: *Probability distributions (blue) and survival functions (orange) for (a) pre-symptomatic period and infectious duration of asymptomatic cases (b) Modelled distribution for the duration of infectivity in mildly infectious cases; and (c) convolution of the distributions in (a) and (b) to produce the distribution of asymptomatic infectious period.*

14. Following exposure, a fraction of exposed individuals remains asymptomatic. Possible values are based on estimates from the United States Centers for Disease Control and Prevention (CDC) [19]:
    - In the best case scenario, 10% of cases are asymptomatic;
    - In the worst case scenario, 70% of cases are asymptomatic; and
    - In the most likely scenario, 40% of cases are asymptomatic.

    From these estimates, the probability distribution of the fraction of infections that are asymptomatic is modelled as a symmetric triangular distribution centred at the most likely scenario.

15. For symptomatic cases, there is a pre-symptomatic period. Empirical results from [20] suggest that the onset for the presentation of symptoms follows a lognormal distribution with

$\mu_p = 1.621$ and $\sigma_p = 0.418$. (This distribution gives a median period of 5.1 days.) The survival function for this function is,

$$S_p(t) = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{\ln t - \mu_p}{\sqrt{2}\sigma_p^2}\right), \tag{6}$$

where $\mathrm{erf}(\cdot)$ is the error function. Once symptomatic, individuals self-identify and are not a concern for sentinel testing.

16. The duration of asymptomatic cases is more difficult to determine, particularly given that the delineation between mild infections and asymptomatic infections is not always clear. For example, some add an additional type of infection, pauci-symptomatic, where symptoms are present but too subtle for the individual to think anything of them. While pauci-symptomatic cases are not considered here, due to the fuzzy boundary between symptomatic and asymptomatic cases and following [11], a conservative estimate for the duration of asymptomatic infections is taken as the time it takes for a symptomatic case to present symptoms *plus* the time that mild symptomatic cases are infectious. The latter is derived from [21], who related their ability to grow an infectious culture of the virus from swabs to the likelihood of an individual still being infectious. From there, they tie the probability of being infectious to the time since symptom onset. Fitting the data from [21] to the survival function of their results suggests a gamma function with parameters $\mu_g = 4.6$ d and $\sigma_g = 1.7$ d. The survival function is described as,

$$S_i = S_i(t) = 1 - \frac{1}{\Gamma(\kappa)}\gamma\left(\kappa, \frac{t}{\theta}\right) \tag{7}$$

where $\kappa = \mu_g^2/\sigma_g^2$, $\theta = \sigma_g^2/\mu_g$, $\Gamma(\cdot)$ is the gamma function and $\gamma(\cdot)$ is the incomplete gamma function.

The survival function for asymptomatic cases is then,

$$S_a = S_a(t) = S_p(t) * S_i(t), \tag{8}$$

where $*$ is the convolution operator. Figure 3 shows the survival functions S_s, S_i, and S_a, as well as their associated probability density functions, in panels a, b, and c, respectively.

17. The survival function defining the limit of when testing will be effective is written as,

$$S = S(t) = p_a S_a(t) + (1 - p_a) S_s(t), \tag{9}$$

where p_a is the probability that a given case will be asymptomatic. Where one assumes uncertainty in p_a and its value is defined by a probability distribution $p(p_a)$, Equation (9) becomes,

$$S = S(t) = \int_0^1 p(p'_a)\left[p'_a S_a + (1 - p'_a) S_s\right] dp'_a. \tag{10}$$

18. Following Equation (10) we estimate the survival function for all infections to be,

$$S_c = S_c(t) = \int_0^1 p(p'_a)\left[p'_a S_a + (1 - p'_a)(S_p * S_{si})\right] dp'_a. \tag{11}$$

where S_{si} is the survival function for the symptomatic phase of symptomatic cases, estimated using a gamma distribution with $\mu_s = 10$ d and $\sigma_s = 3$ d.

19. Each individual has equal likelihood of infection, and a single infection in the group does not change the likelihood of infection outside of the workplace for other individuals.

3 Freedom from Disease (FFD)

The starting point for this work stems from methods used in veterinary medicine to test if a population of livestock in a large region (e.g., a country) is free of a particular disease [9, 22]. Foddai et al. [8] recently proposed the application of these strategies to the problem of assessing whether COVID-19 is present on a base, citing Epitools [23] as a resource for making the necessary calculations. The text below describes the concept of FFD as it is used on the Epitools website [24],[4] then describes its mechanics and limitations for the scenario shown in Figure 1.

For a population to be truly free from disease, every member of the population must not be infected and to ensure this, every member of that population needs to be tested with a perfect test that will not report any false positives or false negatives. Such tests do not exist for identifying a SARS-Cov-2 infection. However, even *if* such a test existed and *if* it were possible to test everyone, another problem arises: for a subpopulation mixing with a larger population in which the disease exists, it is possible (if not likely) that someone who initially tests negative will become infected later. This results in a need for constantly testing everyone, which is not realistic.

The FFD sentinel approach uses a series of random tests where, under the assumption of no false positives, sentinel testing is carried out where, for each round of testing, k individuals are selected randomly from the subpopulation.[5] If tests do come back positive, then the disease is present in the group. The sampling strategy is thus aimed at providing a confidence that the point prevalence is below some background level, or *design prevalence* [10]. The model applies a Bayesian approach where previous confidence of freedom is increased by each tranche of testing. However, following testing, the confidence drops because there is a probability of the virus being introduced to the group over time (Figure 2). For the approach to be successful, the confidence gained by each tranche of testing must be higher than or equal to the amount lost due to the probability of introduction of the virus to the group between tranches, with results gravitating towards an acceptably high confidence level. (For example, one may want to have 95% confidence that the point prevalence in the subpopulation is at or below 0.5%.) Thus, if the prevalence and desired confidence levels are known, one can design a sampling strategy to reach that confidence level.

3.1 Mathematical description

3.1.1 Probability of being free from disease

This section provides a mathematical formulation for FFD. To begin, we assume a prior belief for the probability that the subpopulation is disease-free, i.e., that $p(\pi = 0)$. Following a sequence of tests, this belief can be updated using Bayes theorem:

$$p(\pi = 0|T^+ = 0) = \frac{p(T^+ = 0|\pi = 0)p(\pi = 0)}{p(T^+ = 0|\pi = 0)p(\pi = 0) + p(T^+ = 0|\pi > 0)p(\pi > 0)}. \tag{12}$$

[4] This information was not present on the website, but required significant review of the open literature.

[5] Generally, it is assumed that false positives can be ruled out through means such as isolating and retesting, i.e., in veterinary medicine removing an animal from the herd until the positive test is overturned [25]. This is a less-than-ideal scenario when examining the situation for COVID-19 due to the mixing rate of the population of interest with the background population. Too many false positives could send many people into self-isolation, and the timescales for retesting may be slower than the timescale for making decisions about a potential outbreak.

Under the assumption of perfect test specificity, $p(T^+ = 0|\pi = 0)$ is unity, so this can be simplified to

$$p(\pi = 0|T^+ = 0) = \frac{p(\pi = 0)}{p(\pi = 0) + p(T^+ = 0|\pi > 0)[1 - p(\pi = 0)]}. \tag{13}$$

3.1.2 Design prevalence and customizing the sample size

An aspect of FFD is that, if a specific confidence level is desired, one can adjust the number of tests in the next round of sampling. In Equation (13), the term $p(T^+ = 0|\pi > 0)$ can be rewritten as $1 - p(T^+ > 0|\pi > 0) = 1 - Se_s$, where Se_s is the sensitivity of the sentinel strategy, i.e., the probability that the testing strategy will return a positive test if there is a case in the population. Equation (13) can thus be rearranged to give a relationship between the prior, $p(\pi = 0)$, the desired posterior, $p(\pi = 0|T^+ = 0)$, and the sentinel sensitivity required to reach that desired posterior:

$$Se_s = 1 - \frac{p(\pi = 0)}{1 - p(\pi = 0)} \left[\frac{1}{p(\pi = 0|T^+ = 0)} - 1 \right]. \tag{14}$$

For example, if the prior probability is 0.8 and the desired posterior probability is 0.9, then the sentinel sensitivity must be $1 - 0.8/0.2(1/0.9 - 1) = 0.\dot{5}$.

To customize the sample size, we must next determine how many individuals must be tested to achieve the Se_s-value prescribed by Equation (14). Assuming perfect specificity, if there is one infected individual in the group, then the probability that the individual in the group will test positive is $Se\, k/N$.

To move to the step of specifying k to reach a certain confidence of FFD, we must first introduce a concept known as "design prevalence," denoted as π_d. This value is needed to reject the null hypothesis that the disease is present in the group [10], and it comes into play because the higher/lower the prevalence, the fewer/more the number of tests required to detect the presence of the virus. For COVID-19 infections in our subpopulation, we set the design prevalence to the background point prevalence for the population at large.[6] Based on our design prevalence, the expected number of individuals in the subpopulation who would be infected is $d^* = N\pi_d$ (rounded to the nearest integer) and so the probability of nobody in the group testing positive is $(1 - Se\, k/N)^{d^*}$. Since the sentinel sensitivity is the probability of getting a positive test,[7]

$$Se_s = 1 - \left(1 - \frac{Se\, k}{N}\right)^{d^*}. \tag{15}$$

Hence, to identify the number of tests required to achieve a desired sentinel sensitivity, we rearrange Equation (15) to get,

$$k = \left\lceil \frac{N}{Se} \left[1 - (1 - Se_s)^{1/d^*} \right] \right\rceil, \tag{16}$$

where $\lceil \cdot \rceil$ is the ceiling operator (e.g., $\lceil 7.3 \rceil = 8$).

[6] If additional information about the group of interest relative the average of the population is known, then this information can also factor into the value used for the design prevalence.

[7] This is an approximation that speeds up the computation considerably and holds for a large populations where $N \approx N - d^*$. In cases where this approximation begins to break down the value can be estimated iteratively using a hypergeometric distribution. The approach would be to evaluate Equation (2) for increasing k values with $d = d^*$ and $x = 0$ until $p(T^+ = 0|N, k, d^*, Se, Sp = 1) \leq (1 - Se_S)$.

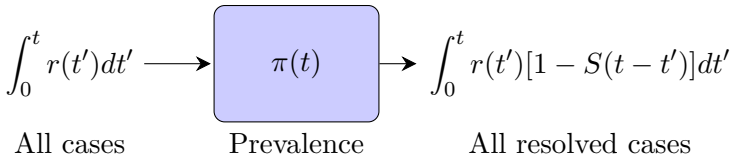


Figure 4: *Box model used to relate prevalence and incidence rate.*

3.1.3 The probability of introduction and discounting the prior

At the onset of the first round of testing, one must make an estimate for the prior probability of an infection within it. (Note that starting with a prior of 0 is not a valid starting point as it does not allow the posterior to move away from that value. A reasonable approach could be to apply a Bernoulli trial to estimate the probability of no infections in the group based on the background or design prevalence.) After the first round of testing, the probability of an infection in the group shifts to the posterior probability following Equation (13).

Now, consider a situation where testing is conducted at regular intervals τ (e.g., if testing were once a week, then $\tau = 7$ d). When the next round of testing begins, we have some information and the new prior is informed by the posterior from the previous round of testing. However, this posterior needs to be discounted due to the probability that the virus may have been introduced to the group since the last round of tests. Considering a regular sentinel testing regime where tests occur on days $\{0, \tau, 2\tau, \ldots\}$, the prior at time $(n+1)\tau$ is derived from the posterior at time $n\tau$ as,

$$p_{n+1}(\pi = 0) = p_n(\pi = 0|T^+ = 0)\left[1 - p\left(I|\tau\right)\right], \tag{17}$$

where $p(I|\tau)$ is the probability that an infection will have been introduced to the group between tests separated by the time period τ.[8]

Determining the probability of introduction to the group depends on the rate of new infections, or incidence rate r, within the region. The relationship between r and π depends on how quickly cases resolve (Figure 4), and for constant π, the incidence rate is equal to the rate at which cases resolve. Assuming that initially at $t' = 0$ the prevalence is zero, then the point prevalence a time t is defined as the sum of all infections that have not resolved,

$$\pi(t) = \int_0^t r(t')S(t - t')dt', \tag{18}$$

where S is the survival function for the duration of infections (that is to say $S(\Delta t)$ is the fraction of infections that will not have resolved at a time Δt after the onset of infection).

For constant prevalence, the relationship above simplifies to

$$\pi = r\int_0^\infty S(t')dt' = r\mu_i, \tag{19}$$

where μ_i denotes the average duration of an infection.

The probability of introduction to the group, $p(I|\tau)$, can be estimated by taking the incidence rate r and noting that, since π is the probability that any individual selected randomly from the

[8] If the point prevalence is constant, then $p(I|\tau)$ is constant, but if it is changing in time, then this term too will have time dependence. Considering the situation for time-dependent prevalence is beyond the scope of this effort but could be adapted from methods in [16].

population may be infected, r is also the daily probability that an individual in the population will contract the disease. For the population of interest, assuming that the probability that one might get infected on any given day can be treated as a Bernoulli trial, the probability that the virus is introduced to the group after a time τ is,[9]

$$p(I|\tau, r, N) = 1 - (1 - r)^{N\tau}. \tag{20}$$

3.2 Application

Due to the mixing rates with the general population, which are much harder to control with people than livestock, it may never be possible to get a high degree of confidence that the population is free from disease. In fact, for sufficiently low design prevalence and an imperfect test, applying Equation (16) indicates that it may be impossible to reach an adequate degree of confidence that the group is free from disease.

Based on current knowledge of COVID-19, it is possible to relate prevalence to the fraction of the population that is likely to be infected each day and, from that, one can determine the likelihood that somebody within the group of interest becomes infected over a given time period. Using $r = \pi/\mu$ from Equation (19) and applying Equation (20), Figure 5 shows the expected probability of introduction to a group based on the background point prevalence.

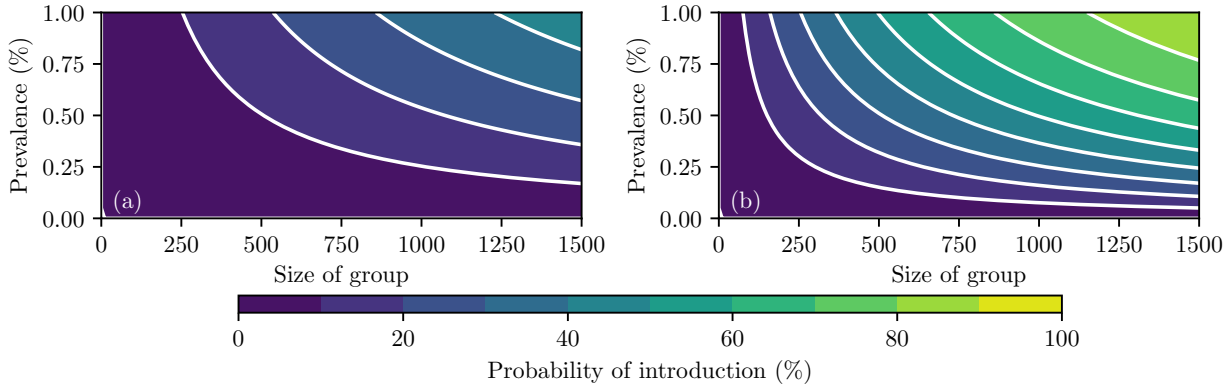
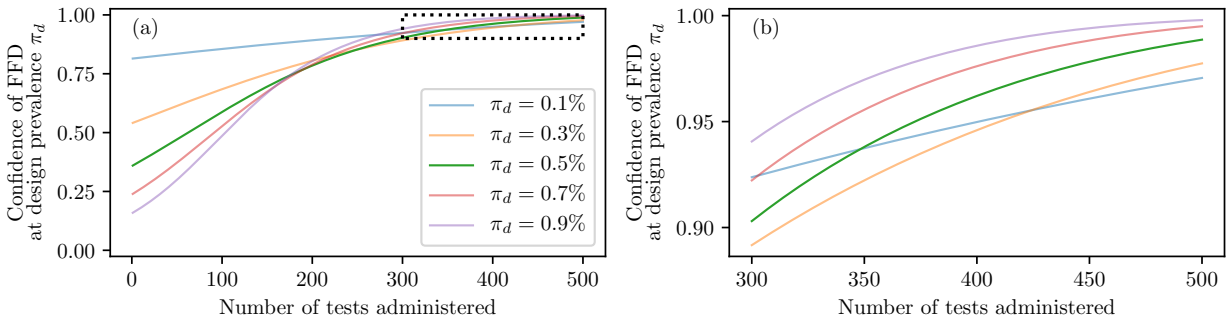


Figure 5: *Relationship between group size, background point prevalence, and (a) daily or (b) weekly probability of introduction to the group.*

The probability of introduction makes it very difficult to test at a level and frequency that can assure FFD. For a building of $N = 500$, Figure 6 shows the relationship between the number of tests per week and the confidence of being free from disease for different design prevalence values ranging from 0.1% to 0.9%. When the number of tests is low, the confidence in freedom from disease (at the design prevalence for the model) is higher for low values of design prevalence simply because, due to the laws of probability, the group is more likely to be free from disease (Figure 6a). As testing increases, the confidence increases for the high prevalence curves faster than the low prevalence ones because it is more unlikely to have that many negative tests at high design prevalence (Figure 6b).

[9] Methods for estimating $p(I)$ with time-dependent point prevalence would require looking at the estimate for r each day (or at time step):

$$p(I|\tau, r, N) = 1 - \prod_{t=n\tau+1}^{(n+1)\tau} (1 - r(t))^N.$$



0.8

Figure 6: *Confidence in FFD after one set of tests, where the initial prior is set based on the expected density of cases within a group of $N = 500$ individuals for the design prevalence values stated. Tranches of tests are administered after one week of mixing with the general population. Results assume a test with sensitivity $Se = 0.7$, and the lines show results for different design prevalence estimates. Panel b shows a close-up of the dotted black region in Panel a.*

Based on these results and the fact that false positives may be difficult to rule out in a timely manner, it may not be possible to confidently demonstrate FFD for a group of people that is mixing regularly with the general population. For a group of $N = 500$ individuals and a prevalence of $\pi_0 = 0.5\%$, the weekly probability of introduction to the group is $p(I|\tau = 7) = 27\%$. For a test with a 30% false negative rate (i.e., a sensitivity of $Se = 0.7$), maintaining that the building is free from disease with 95% confidence given a design prevalence of our assumed background population (i.e., $\pi_d = \pi_0 = 0.5\%$) would require testing 360 individuals each week (72% of the building) with all tests coming back negative. If only 10% of the subpopulation is tested each week, then the confidence that the group is FFD at the design prevalence is 51% (Table 2).

Table 2: *Relationship between weekly testing volume and FFD confidence for different design prevalence values.*

Design prevalence π_d	Probability of introduction $p(I)$ Daily ($\tau = 1$)	Weekly ($\tau = 7$)	Number of tests for 95% FFD confidence	FFD confidence with 50 tests
0.1%	1.9%	6.1%	378	85%
0.3%	5.5%	17%	396	65%
0.5%	9.0%	27%	360	51%
0.7%	12%	36%	329	42%
0.9%	16%	43%	303	34%

Of further concern for this example is the fact that test results are not instantaneous. Assuming a one-day delay from testing to test results, then for $\pi_0 = 0.5\%$ this delay would result in a 9% discount in the probability of being FFD. Hence, to have 95% confidence upon release of test results is no longer possible, since having 100% confidence would have been discounted to 91% the following day. Using the example of 50 tests per week, the initial 51% confidence described above would be reduced to 46% the if reporting of test results occurs the day after testing.

In summary, this approach may pose considerable difficulties when trying to consider a large population during COVID-19. In the situation where the mixing of the population cannot be regulated, the number of tests required may become infeasible (i.e., more tests than practical) or even impossible (i.e., more tests than the number of people in the subpopulation). In reality, tests do not have perfect specificity, and decisions may be necessary prior to the ruling out of false positives (see Footnote 5). In fact, for larger groups mixing with the population, the rate of introduction might be sufficiently high that the value of an overturned false positive may be negated by the time the false positive is recognized. This suggests that it is necessary to incorporate false positives into the analysis, which is done in Section 4. Last, if one assumes that symptomatic individuals will self-identify and that detecting asymptomatic individuals is only useful if it is done before they are no longer infectious, then sentinel testing is limited in that it must identify infected individuals who are either asymptomatic and still contagious or have not yet developed symptoms. This issue has is not addressed above at all, but is discussed in Section 5.

4 Extending the FFD concept to account for true and false positive tests

Demonstrating FFD is ideal, but based on the analysis above it is not clear whether this is possible for the problem at hand given disease prevalence and the probability of introduction of the virus in time. Above, we showed that in a situation with $N = 500$ and a point prevalence of $\pi_0 = 0.5\%$, it is difficult to ensure that FFD unless virtually the entire population is tested (e.g., Table 2). In addition, the possibility of false positives adds to the challenge of making decisions using FFD results in a timely manner since, by the time a false positive is ruled out, the results from that round of testing may be significantly discounted due to mixing.

In this section, we next explore whether an alternate approach could be to design a sentinel sampling strategy that, instead of aiming to demonstrate FFD, aims to catch an outbreak early. The premise of the approach is that, if one accepts that there may be a few infections within the group due to building occupants being part of with their community (Figure 1), the challenge then shifts from trying to ensure that the virus is not in the building to instead trying to ensure that the building environment itself is not causing a source of spreading, i.e., that the building and its occupants do not create the conditions for an outbreak. The premise becomes that, so long as the prevalence in the building is not above the background point prevalence, the work environment is unlikely to be a source of spreading, and the question becomes whether this alternative is feasible.

When looking at the possibility of an outbreak, the focus is very similar to FFD except that now the effort is to show the probability of being at below the community point prevalence. Two major differences between the FFD model and this model are that the approach below incorporates the implications of positive test results and considers non-perfect specificity (i.e., false positives are possible).

In addition to these inclusions, uncertainty in the point prevalence is also considered while extending the FFD framework. This consideration could, in principle, be used for FFD as well but that extension is not considered here. We include two sources of uncertainty in the point prevalence estimate. First, when looking at the point prevalence in a given region, there will be some uncertainty about that value [7] so instead of using a single point prevalence value, we consider a distribution, which for this effort is modelled simply as a beta distribution. Second, when drawing the subpopulation of the building out of a larger population, one is conducting a Bernoulli trial in which the resulting prevalence value could be above or below the background population, simply by chance. Both additions lead to the same end result where the point prevalence in the group is considered as a distribution rather than a single value.

4.1 Mathematical description

4.1.1 Interpreting test results without considering background point prevalence

The sentinel sampling strategies discussed here employ a Bayesian approach that considers the background point prevalence. To outline the need to consider the background point prevalence in interpreting test results, consider the challenge of interpreting test results without it. Figure 7 shows the probability distributions for the point prevalence in the group based on test results. The figure shows the probability distributions of the background point prevalence when $k = 50$ out of $N = 500$ individuals are randomly tested. Panel a considers the case with perfect specificity and

Panel b considers the case with an assigned specificity of $Sp = 0.99$. If one does not use beliefs about the background point prevalence then it can be extremely difficult to interpret test results. For example, if a single test comes back positive, it is difficult to say with any certainty whether the point prevalence is 0% or 5%, particularly for the case of imperfect specificity (Figure 7b). Below, as the different methods are discussed, it will be demonstrated how consideration of the background conditions, when combined with testing results, can improve result interpretation.

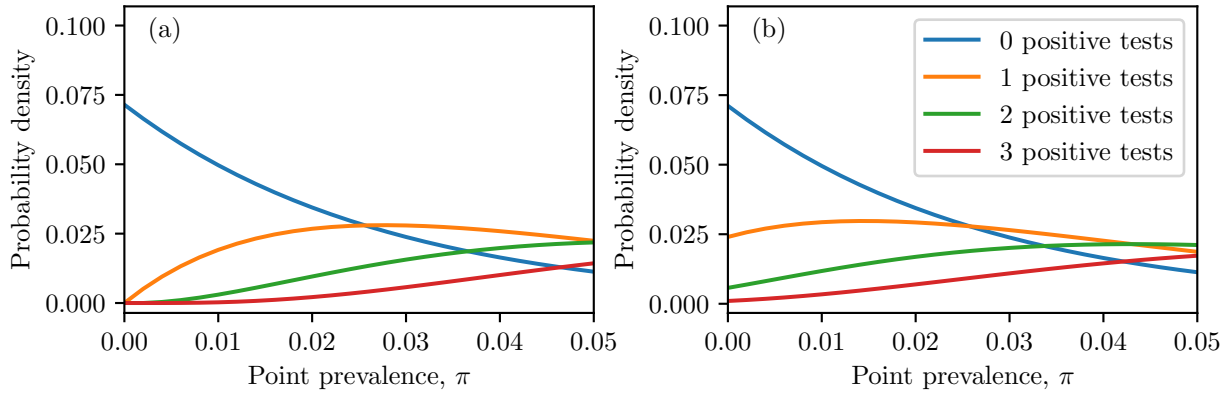


Figure 7: *Probability distributions of prevalence based on test results for (a) $Sp = 1$ and $Se = 0.7$; and (b) $Sp = 0.99$ and $Se = 0.7$.*

4.1.2 A Bayesian estimation of point prevalence

Assume for a moment that we believe the point prevalence in the region fits some distribution $p(\pi_0)$. Armed with this belief, we can revisit results above using Bayes rule. Given the prior belief regarding the prevalence π_0, the posterior probability that prevalence is π given x positive tests is,

$$p\left(\pi|T^+ = x\right) = \frac{p\left(T^+ = x|N,k,d,Se,Sp\right)p\left(\pi = d/N\right)}{\sum_{d=0}^{N} p\left(T^+ = x|N,k,d,Se,Sp\right)p\left(\pi = d/N\right)}. \tag{21}$$

Initially, the prior probability for π is based on the background point prevalence, but is updated to the posterior of Equation (21). Just as there was a probability of introduction of the virus above, the same dynamics cause the posterior point prevalence to relax in time. However, since we are not trying to demonstrate FFD but rather that the local point prevalence of the subpopulation of interest is at or below regional point prevalence, the distribution relaxes back to the regional point prevalence over time. The approach for determining this relaxation is outlined below.

4.1.3 Discounting point prevalence estimates in time

Discounting point prevalence estimates in time builds on the approach used to estimate the probability of introduction in the methods defining the FFD approach. Consider the situation where, after testing, we assume a background point prevalence in the subpopulation to be $\hat{\pi}$, while our understanding of the regional point prevalence may suggest a different value π_0.[10] Here, we treat the

[10] The background point prevalence need not be *exactly* the regional point prevalence. For example, one could account for demographics of the workforce or other factors to presuppose a background value that is some modification of the regional point prevalence.

problem as one where the incidence rate prior to testing was constant and achieved a steady point prevalence value $\hat{\pi}$, meaning that $r = \hat{\pi}/\mu$, as per Equation (19). However, given our knowledge of the background point prevalence, then similar to the probability of introduction, the incidence rate following testing relates to the background point prevalence $r = \pi_0/\mu$. Hence, if we consider a situation where sentinel testing at a time t_0 suggests a point prevalence $\hat{\pi}$ but where we are now estimating the point prevalence at some time $t > t_0$, extending Equation (18) for the current problem gives,

$$\pi(t > t_0) = \frac{1}{\mu}\left[\hat{\pi}\int_0^{t_0} S(t-t')dt' + \pi_0\int_{t_0}^{t} S(t-t')dt'\right]. \tag{22}$$

Applying this result to probability distributions for the local and background point prevalence after sampling gives

$$p(\pi|t \geq t_0) = \frac{1}{\mu}\left[p(\hat{\pi})\int_0^{t_0} S(t-t')dt' + p(\pi_0)\int_{t_0}^{t} S(t-t')dt'\right], \tag{23}$$

where $p(\hat{\pi})$ and $p(\pi_0)$ are the probability distributions for $\hat{\pi}$ and π_0, respectively.

In the situation where the group of interest is being more careful than the general population, the point prevalence in the group is likely below the background point prevalence so relaxing to the background point prevalence is a cautious approach. When point prevalence in the group is likely above background, however, relaxing to the background may provide an incorrect suggestion that the point prevalence in the group drops after testing. While there is a possibility that there are extra cases by chance, in which case relaxing to a lower value may be justified, it is also possible that the work environment is contributing to the spread and sentinel testing alone cannot identify which case is taking place. Once the virus is spreading in the workplace and an outbreak is taking place, the sentinel sampling strategy would provide little utility for monitoring. The approach, therefore, requires the perceived point prevalence in the group to be below the background point prevalence, and should not be used once the perceived group point prevalence raises above background levels.

4.2 Application

Instead of trying to determine that the group is FFD, we now attempt to estimate the likelihood that the point prevalence in the group matches what we would expect based on the background point prevalence. Again, we consider a building with $N = 500$ individuals but now assume a point prevalence $\pi = 0.5 \pm 0.125$, fit to a beta distribution, and estimate the prevalence in our group of individuals based this estimate of the regional point prevalence and the fact that we are drawing N people randomly from it.[11] Figure 8 shows the resultant distributions for the general population (panel a) and for the subpopulation (panel b). The discrete values shown in panel b are due to the fact that there are only $N = 500$ individuals in the subpopulation, so there are only 501 possible values for the point prevalence in the group. Only the first 11 are shown, since the values are almost zero beyond $\pi = 2\%$, which would be equivalent to 10 infections in the group of 500.

[11] For a mean μ and a standard deviation σ, the parameters for a beta distribution are $\alpha = \mu/\sigma$ and $\beta = (1-\mu)/\sigma$. Drawing 500 samples randomly from the beta distribution in Figure 8a gives the beta-binomial distribution shown in Figure 8b.

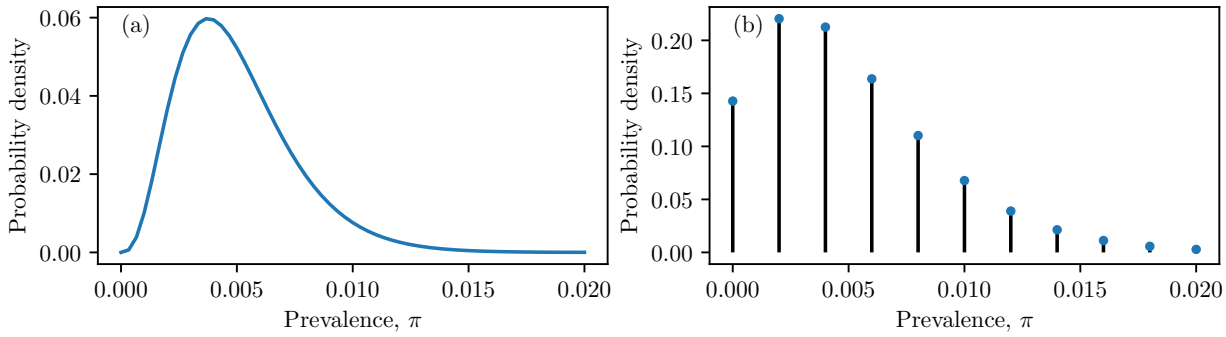


Figure 8: *(a) Distribution of point prevalence (assuming a beta distribution) with $\pi = 0.5\% \pm 0.125\%$. (b) Distribution of point prevalence within a group of $N = 500$ individuals assuming random sampling from the distribution in (a).*

For this model, we consider two examples:

- Perfect specificity and 70% sensitivity ($Sp = 1$ and $Se = 0.7$); and
- 99% specificity and 70% sensitivity ($Sp = 0.99$ and $Se = 0.7$).

The purpose of these examples is to show how a test with imperfect specificity affects the outcome for the model. Comparing the examples underlines the importance of being able to rule out false positives, if possible, and of applying caution in interpreting results when one cannot rule out false positives.

As noted in Section 4.1.1, if one does not use beliefs about the background point prevalence then it can be extremely difficult to interpret test results (Figure 7). Accounting for the background distribution (Figure 8b) and applying Equation (23) helps to recast the results and improve interpretability (Figure 9). In this figure, the grey filled distribution shows the expected background distribution (prior) and the lines show the posterior distribution based on the likelihoods shown in Figure 7 and the background distribution.

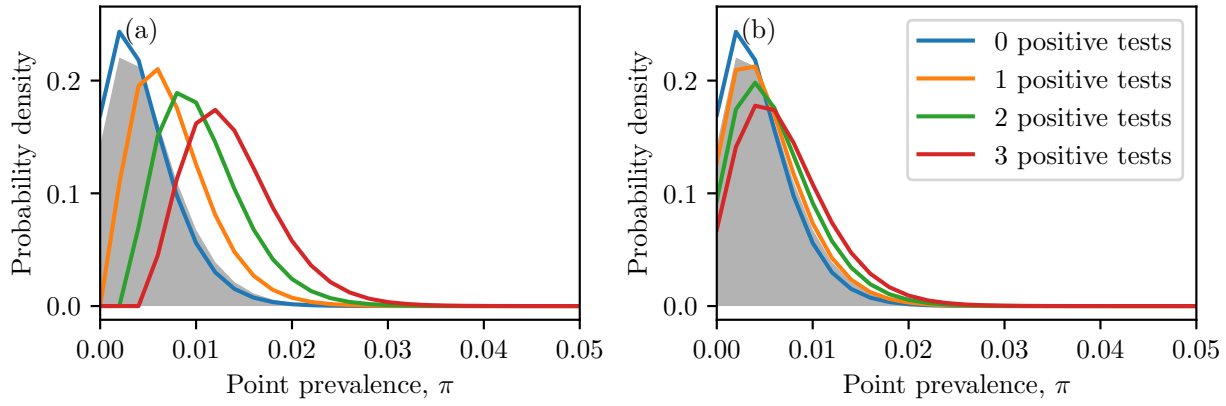


Figure 9: *Probability distributions of point prevalence based on test results and background belief that point prevalence matches Figure 8b. (a) $Sp = 1$ and $Se = 0.7$; and (b) $Sp = 0.99$ and $Se = 0.7$.*

Comparing the two panels of Figure 9 shows the effect of imperfect specificity on interpreting the results. In the case with perfect/imperfect specificity, the outcomes for different numbers of tests are more separated/clustered. If one characterizes an outbreak as a situation where the point prevalence for the group is above the expected background point prevalence, we can quantify the effect of this separation/clustering of results by estimating the probability of an outbreak in the group.[12] The premise of this approach is that, if testing suggests cases are above what would be expected given the background point prevalence, there is an increased likelihood that the group is experiencing an outbreak and further investigation is recommended.

The resultant probability of being above background levels, for the perfect and imperfect specificity examples are presented in Figure 10 and Table 3. In the best case scenario of no positive tests, the likelihood of an outbreak sits at 38%. For perfect specificity, the likelihood increases dramatically as tests increase, growing to 90% by the time three tests come back positive. However, interpretation is much more difficult for the case with a 99% specificity. While the outlook is the same for no positive tests, at three positive tests the likelihood of an outbreak is 56% (rather than 90% for perfect specificity) and five positive tests still gives only a 68% probability of an outbreak. This result may give pause to consider how one might interpret a set of positive test results, particularly if one assumes the test has imperfect sensitivity.

Table 3: *Probability of an outbreak based on test results.*

Test	Number of positive tests					
specificity	0	1	2	3	4	5
$Sp = 0.99$	38%	44%	50%	56%	62%	68%
$Sp = 1.00$	38%	62%	80%	90%	96%	98%

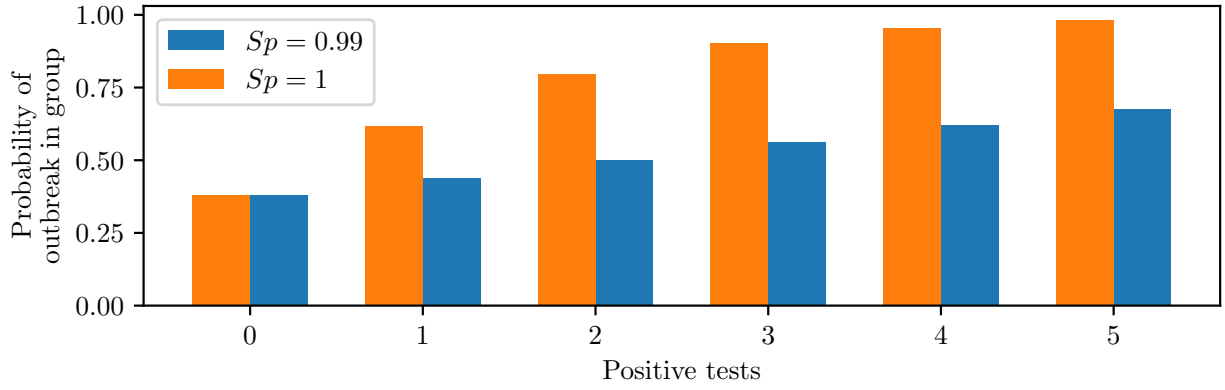


Figure 10: *Probability of an outbreak in the group for* $N = 500$, $\pi = 0.5\% \pm 0.125\%$, *and* $k = 50$ *individuals in the group tested at random.*

[12] This is achieved by looking at the joint distribution of the background point prevalence and the posterior estimate of the prevalence and determining the probability that the latter is greater than the former. Given two distributions X and Y with a joint probability distribution $f(x, y)$ the probability that X is greater than Y is,

$$p(X > Y) = \int_{-\infty}^{\infty} \int_{y}^{\infty} f(x, y) dx dy.$$

4.3 Application to a regular sentinel testing plan

Like the FFD approach above, this extension runs into the same difficulties due to the constant probability of introduction of new cases. However, because we are beginning from an assumption where the disease is already in the group, determining how to alter the distribution in time is slightly more complex.

Figure 11 shows the probability that the population in the group is above background point prevalence given the testing regime where once each week, $k = 50$ of the $N = 500$ individuals are randomly selected for sentinel testing. The figure shows a scenario where tests take place on Days 1, 8, 15, and 22. In the initial weeks, all tests come back negative but various outcomes are presented for the testing on Day 22. Initially a saw-tooth pattern is present because, while getting no positive tests decreases the likelihood of being above background point prevalence, the constant probability of introduction slowly raises the risk back up to the levels one would expect if no testing were taking place. After the next two subsequent sets of tests, the risk decreases again as we have assumed no positive tests. After each of those sets of tests, the probability again relaxes back towards the background level. However, the possible outcomes diverge for the fourth test on Day 22, with the different curves inferring different probabilities of being above the background point prevalence based on different numbers of positive tests. Note that, if the probability of being above background is now greater than what it would have been without testing, the curves no longer relax back towards that background expectation. At this point, if the prevalence in the building is above background, then there is a chance the building environment is contributing to the spread and it makes little sense to relax back to a state where that is not the case. Panel a shows the result for a test for perfect specificity and Panel b shows the result for a test with $Sp = 0.99$. In both cases, the sensitivity is $Se = 0.7$.

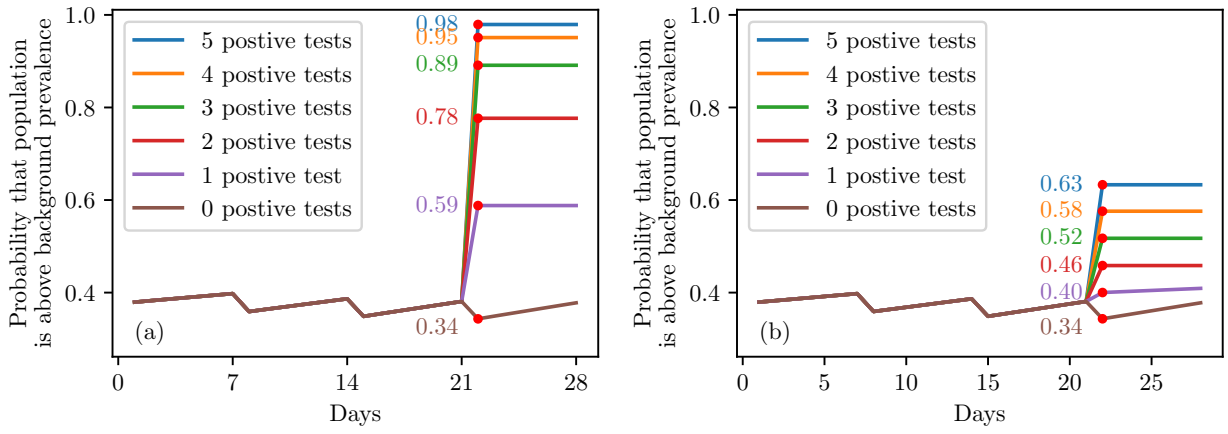


Figure 11: *Probability that the group is at or below background point prevalence for $N = 500$, $\pi = 0.5\% \pm 0.125\%$, and $k = 50$ individuals in the group tested at random each week given (a) perfect specificity and (b) 99% specificity.*

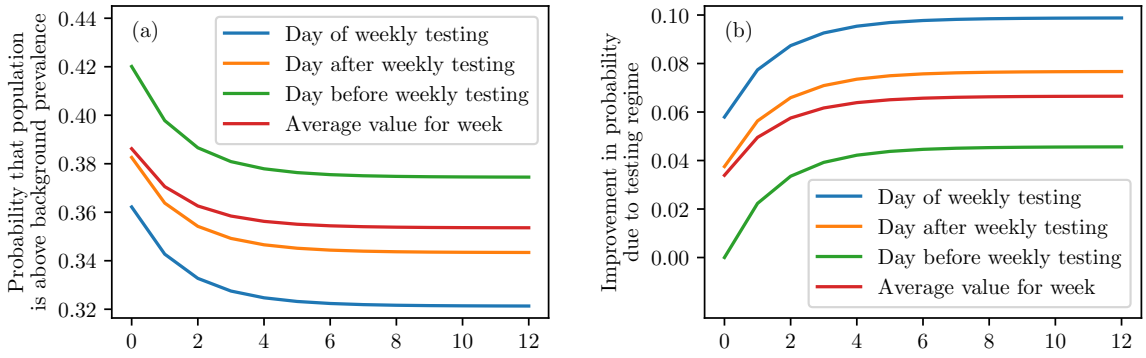


Figure 12: *Evolution of the probability that the group is at or below background point prevalence, assuming no positive tests, for* $N = 500$, $\pi = 0.5\% \pm 0.125\%$, *and* $k = 50$ *individuals in the group tested at random each week given. Panel a shows the probability values while Panel b shows the improvement in probability compared to the scenario with no testing.*

Like Table 3 and Figure 10, Figure 11 again presents the challenge of interpreting results given imperfect test specificity. However, careful inspection reveals that the numbers shown after testing in Figure 11 are slightly below those shown in Table 3. This difference is because the saw tooth pattern seen in the figure has a slight downward trend, due to the time required for the testing strategy to equilibrate to a state where the weekly increase balances the weekly drawdown due to testing. The time required to reach this equilibrium state is shown in Figure 12, which shows time series for the state of knowledge before and after testing, the state of knowledge assuming a one-day delay between testing and getting results, and finally an average probability for the week.[13]

While Figure 12 may suggest that testing (in the absence of positive tests) could increase certainty that prevalence in the group is below the background level, as soon as one gets a positive test, the difficulty of interpreting the results given imperfect specificity will enhance the difficulty in determining how to act (Figures 10 and 11, Table 3).

One may choose to retest those who test positive, but by the time the second set of test results come back the values from the first test will have discounted considerably. In addition, one would isolate those who did test positive and initiate a contact tracing effort in order to reduce the risk of the virus spreading. Omitting these actions would compromise the reason for sentinel testing in the first place. At the time of writing, the test specificities for SARS-Cov-2 tests is excellent, but this may change with the advancement of rapid testing technologies the application of pooled testing and as such it is an important factor to consider.

[13] In Figure 12a, the top curve initially shows the probability of being above background point prevalence in the absence of testing. The value sits at about 42%, rather than 50%. The reason for this is because, when comparing the discrete distribution of background point prevalence (Figure 8b) to itself, there is an approximately 16% chance that the values drawn from the distributions match (using the approach described in Footnote 12) and a 42% chance that the draw from one of the distributions will be above or below a draw from the other.

5 The probability of detecting an unknown infection in a timely manner

The sections above provide a strong argument that it is unlikely that sentinel testing will provide an adequate approach for ensuring FFD (Section 3) or providing much confidence that the group is at or below the regional background point prevalence (Section 4). However, one may still be inclined to use sentinel testing in hopes of catching some cases before they have sufficient time to spread the disease. This section considers that problem, estimating the likelihood of testing and detecting an infected individual before that person either presents symptoms if the case is symptomatic or ceases being a risk to the group because she or he is no longer contagious if the case is asymptomatic. In fact, as we find here, a weekly testing regime is still likely to miss a considerable number of cases, even if everyone is tested each week using a test with perfect sensitivity.

The reason for this lowered performance compared to what one might expect is that there is a time-window in which testing must be done. It is not enough to test an individual who is infected, but they must be tested after they have a sufficient viral load for the test to work and before it is too late for the test to provide useful information to protect the group. If a symptomatic case has evolved to the point where the individual has developed symptoms, then we expect the person to at least isolate themselves from the workplace and hopefully also to get themselves tested. Similarly, if an asymptomatic individual has cleared the virus sufficiently that they are no longer infectious, then there is decreased benefit to identifying them through sentinel testing as they are no longer a risk.

5.1 Defining the time window for testing

As alluded to above, there are two factors to consider in defining the window: when it opens and when it closes. Here, we assume that the time window opens when there is a sufficient viral load for a test to be effective and closes when symptoms present themselves for symptomatic cases or when the individual is no longer infectious for asymptomatic cases.

With the opening of the test window defined by the cumulative density function $f_v(t)$ (Assumption 13) and the closing of the test window defined by the survival function $S(t)$ (Assumption 17), the test window is defined as,

$$h = h(t) = f_v S, \tag{24}$$

with the symptomatic and asymptomatic components being $h_s = f_v S_s$ and $h_a = f_v S_a$, respectively.

The probability of being inside the time window for testing as a function of time, that is the function h, is presented as the grey filled region in the panels of Figure 13. In each panel, f_v is shown as the blue curve and the survival functions S_s, S_a, and S are shown as the orange, green, and red curves in Panels a, b, and c, respectively.

5.2 Application to sentinel testing

If testing is infrequent, then there will be few opportunities to catch an infection in the test window. If testing is more frequent, then there may be multiple opportunities. If one tests all individuals every three days, and if each of those tests is independent of the previous, then there is a very high

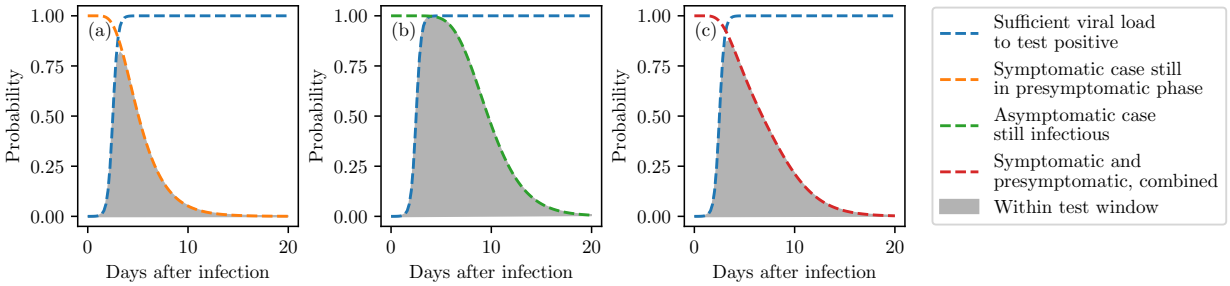


Figure 13: *Window in time during which testing is likely to be effective for (a) symptomatic cases; (b) asymptomatic cases; and (c) all cases.*

likelihood of catching an infection before an infected individual shows symptoms (symptomatic) or stops being infectious (asymptomatic) since each person is likely to be tested after the viral load is sufficient but before the survival function starts to reduce considerably. By contrast, if testing is only once every two weeks, then there is a good chance that the opportunity to catch infections will be missed.

The probability of testing and detecting an individual depends not only on the testing frequency τ but also the fraction of the group tested within each test period. If the probability of becoming infected is constant in time and if an infected individual is tested on the next round of testing following their infection, then the timing of the test could be anywhere on the time interval $[0, \tau)$ after infection with equal probability. Extending to consider multiple rounds of testing, and considering that not everyone is tested within each testing cycle, the probability of being tested within the time window for testing in the m^{th} round of testing since infection is,

$$p_{test} = \frac{p_t}{\tau} \int_{(m-1)\tau}^{m\tau} h(t')dt', \tag{25}$$

where $p_t = k/N$ is the fraction of the population tested in each test period.

Equation (25), however, is not complete as it only describes the probability of being tested if no action follows a positive test. For example, if someone were to test positive during the $(m-1)^{\text{th}}$ testing period, that person would have been removed from the group and would not be part of sentinel testing in the m^{th} period.

To consider past test periods, we first consider a reasonable maximum limit T_{max} after which the test window is effectively 0. Using Figure 13 we set $T_{max} = 20$ d. Then, the number of testing windows that needs to be considered is $T = \lceil T_{max}/\tau \rceil$. Considering past testing windows, then we get,

$$p_{test}(n\tau) = \frac{p_t}{\tau} \int_0^{\tau} \prod_{m=1}^{n-1} \{1 - p_t \, Se \, f_v \left[(m-1)\tau + t\right]\} \, h(n\tau + t)dt, \tag{26}$$

where the term $f_v(m\tau + t)$ identifies the probability that an individual case had sufficient viral load to have tested positive in the m^{th} round of tests since infection. We use f_v in Equation (26) because, due to the nature of the survival function, there is no way that a case that is within the test window could have not been in the test window at an earlier time unless the viral load had been insufficient.

To include test sensitivity, we simply introduce $p_{detect} = Se\, p_{test}$ as the probability of catching an infected individual in a timely manner. Combining the probability of testing in each window then yields the probability of a case being caught as,

$$p_{detect} = Se \sum_{n=1}^{T} p_{test}\,(n\tau)\,. \tag{27}$$

Figure 14 shows the efficacy of each testing window for a single individual (equivalent to $p_t = 1$) with $\tau = 7$ d. Panel (a) shows results using Equation (25) only showing the probability that an individual would be tested in a specific test window, ignoring the fact that if someone tests positive in a given window then they will have been removed from the group and will not be tested in a subsequent round of tests. Panel (b) shows the same results but using Equation (26) with a sensitivity of 1. Panel (c) shows the probability of catching the individual in each time window, assuming a sensitivity of 0.7.

After discussions with subject matter experts [26], it was not clear exactly how the assumed sensitivity of $Se = 0.7$ incorporated the possibility of an initial test missing an infection due to insufficient viral load, although there was suspicion that it was included in the value. Hence, we assume the true value for the probability of catching an unknown infection with sentinel testing to rest between the results of Figures 14b and c, i.e., that for the model in Equation (26), we can treat Se as being between 0.7 and 1. The results suggest that, if someone were to be tested every week, and at some point in the testing regime they became infected, the sentinel tests would have a 33%–47% chance of being positive on the first test, a 11%–13% chance on the second test, and a 0% chance on the third test. Combining these probabilities allows one to calculate the risk of the detection escaping sentinel undetected as 44%–60%, i.e., approximately 50%. It is unlikely, however, that everyone will be tested every test period. Extending the findings in Figure 14, Figure 15 shows the probability of detecting an individual assuming $Se = 1$ (Panel a) and $Se = 0.7$ (Panel b) when different fractions of the population are tested at different time intervals. For the example used in this report of a building with $N = 500$ individuals and a sampling strategy of $k = 50$ the results shown in Figure 15 suggest a 5%–7% chance of detecting them. This suggests that, should an individual become infected, there is a roughly 1-in-17 chance that the individual will be detected by the sentinel strategy of randomly testing 10% of the population each week. The remaining cases will either develop symptoms or remain asymptomatic having not been detected before their infectious period has passed.

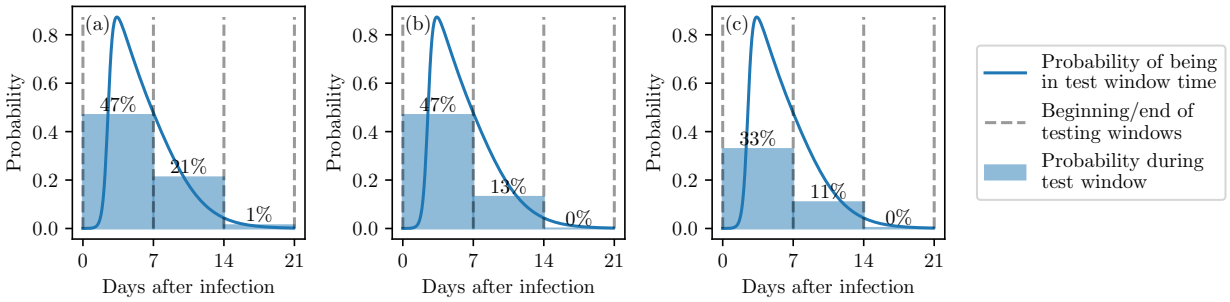


Figure 14: *Expected efficacy of a weekly testing regime for (a) testing a case in a sufficiently timely manner; and (b) yielding a positive test result with $Se = 1$ when inside the test window; and (c) yielding a positive test result when $Se = 0.7$ when inside the test window.*

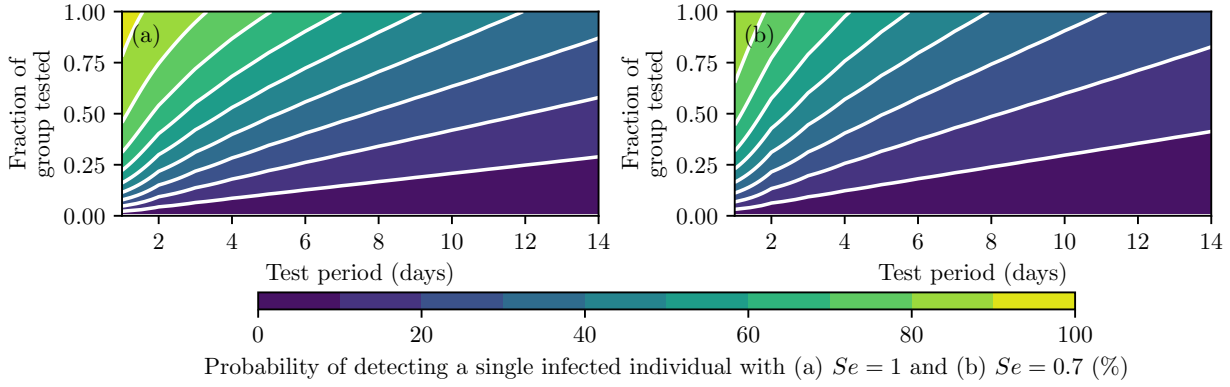


Figure 15: *The probability of detecting a single infection within a group as a function of testing frequency and fraction of group tested.*

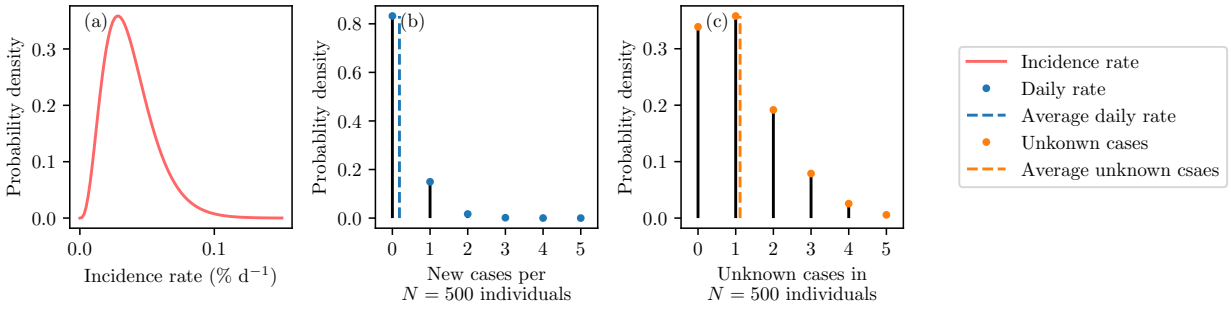


Figure 16: *(a) Incidence rate required to sustain point prevalence; (b) probability distribution for the number of new cases daily in the subpopulation; and (c) the resultant probability distribution of the number of unknown case within the subpopulation.*

5.3 Estimating the number of unknown cases

The section above shows the likelihood of catching a case if it exists, based on the frequency of testing and number of people tested in each round of testing (Figure 15). However, it is still not clear how often this is likely to happen.

Returning to the prevalence model used in the last section (Figure 8), we can use Equation (19) to estimate the rate of new cases as $r = \pi_0/\mu_i$. Applying the relationship between r and π_0 to the distribution presented in Figure 8 gives a probability distribution for the incidence rate r (Figure 16a). The distribution for the arrival rate of new cases within the group r_g is then derived by drawing new cases daily from a binomial distribution (Figure 16b), i.e.,

$$p(r_g|p(r_i), N) = \int_0^1 p(r_i = r') \binom{N}{r_g} r'^{r_g} (1 - r')^{N - r_g} dr', \tag{28}$$

where $p(r_i)$ is the probability density function of the incidence rate r_i (Figure 16a), and the arrival rate of new cases in the group each day must be an integer value, i.e., $r_g \in \{0, 1, \ldots, N\}$. For the scenario considered here ($N = 500$, $k = 50$, $\pi = 0.5\% \pm 0.125\%$), one would expect 0.19 new cases daily, on average.

To estimate the number of infected individuals who could be targeted by sentinel testing, i.e., the number who have an unknown infection, a Monte Carlo simulation was run over 10,000 days where the number of new cases per day was derived from the probability distribution in Equation (28), also shown in Figure 16b, and where each case left the testing window randomly based on the survival function $S(t)$ from Equation (10). The output from this model (ignoring the initial time to ramp up) was then used to generate the probability distribution for the number of unknown cases within the subpopulation at any time (Figure 16c). This result suggests that, an any given time there are very few unknown cases to detect within the population.

5.4 The predictive value of positive tests

The information above allows us to consider how often one could anticipate a true positive test result, relative to the number of positives (i.e., sensitivity of the sampling strategy), as well as the fraction of positive tests that are true positives (i.e., positive predictive value of the sampling strategy).

Combining the probability that an individual who is infected will be tested within a test window (Figure 14) and relating that to the number of people in the subpopulation who carry an unknown infection (Figure 16) allows one to estimate the probability that a positive test in fact reflects an infected individual.

The probability of testing an infected individual can be estimated by first estimating the probability of having an infected individual in your test group given the number of infected individuals in the population of interest, d. (This value can be found using Equation (1).) Applying the resulting value to Equations (26) and ultimately (27) provides an estimate of probability of having a true positive given the number of infected individuals in the group and the size of the group being sampled.[14] The same approach can be used to determine the number of false positives using Equation (4). Finally, performing a weighted average for various values of d using the probability of having d infected individuals in your group (Figure 16c) gives an estimate for the probability of having a true positive or false positive. Table 4 shows the resulting probabilities. Similar to above, two sensitivity values are considered: $Se = 1$ (perfect sensitivity) and $Se = 0.7$.

Table 4: *Probability distributions for positive tests.*

Unknown cases	Probability	Probability of a true positive		Probability of a false positive
		$Se = 0.7$	$Se = 1$	$Sp = 0.99$
0	34.8%	0.0%	0.0%	39.3%
1	36.2%	4.8%	6.8%	39.3%
2	19.0%	9.0%	12.8%	39.2%
3	7.1%	12.8%	18.1%	39.2%
4	2.2%	16.2%	22.7%	39.1%
5	0.7%	19.1%	26.8%	39.0%
6	0.1%	21.8%	30.5%	39.0%
Weighted average		4.9%	6.9%	39.3%

Findings in Table 4 suggest that, given a specificity of $Sp = 0.99$, then for the sampling strategy considered here there would be a roughly 2-in-5 chance of a false positive. That is to say that a false positive could be expected once every two-to-three weeks on average, equivalent to about 20 false positives per year. By contrast, based on the percentages in the bottom row of the table, one would expect a roughly 1-in-20 to 1-in-14 chance of a true positive, equivalent to about 3 to 4 positive tests per year. Assuming one would isolate and contact trace for all positive tests, then essential efforts that must continue will need to plan accordingly. For example, dividing workspaces into split shifts will ensure that if a section of the workplace needs to isolate, there are others available to fulfil the required duties.

[14] While the approach uses the probability of having more than one case in your sample group when $d > 1$ to define p_t, the methods developed for Equation (27) consider only a single infection in the group. Hence, the values presented in Table 4 will slightly underrepresent the true values. However, as the probabilities of having more than one individual in your test group are so small, the weighted values at the bottom of the table are underestimated by less than 2%.

6 Discussion

This report considered the problem of applying random sentinel sampling strategies in hopes of assuring COVID-19 is under control within a population of interest (e.g., an office building) that is mixing with the general population. In Sections 3 and 4, it was shown that the fact that the population of interest is constantly mixing with the broader community makes it very challenging to derive assurances that there are few infections within the group. In Section 3, we consider the FFD model, which requires all tests to come back negative and cannot account for false positives. Even with these restrictions, the constant probability of introduction of the virus to the subpopulation makes it virtually impossible to discern a with high confidence that the population is free from disease.

The findings when considering the time-window for testing in Section 5 suggest that the estimates in Sections 3 and 4 are optimistic, as the sensitivity used in those sections may misrepresent the probability of catching a positive infection. Below, the effect of reducing sensitivity is considered. Additionally, setting specificity to 0.99 affects findings in both Sections 4 and 5. The implications for altering specificity are also considered below.

6.1 Revisiting sensitivity

Figure 17 shows results like those in Figure 6 except with a sensitivity of 0.5, midway through the range proposed in Section 5. In this lower-bound scenario, even if everyone is tested weekly, it is clear that the efficacy of testing is decreased.

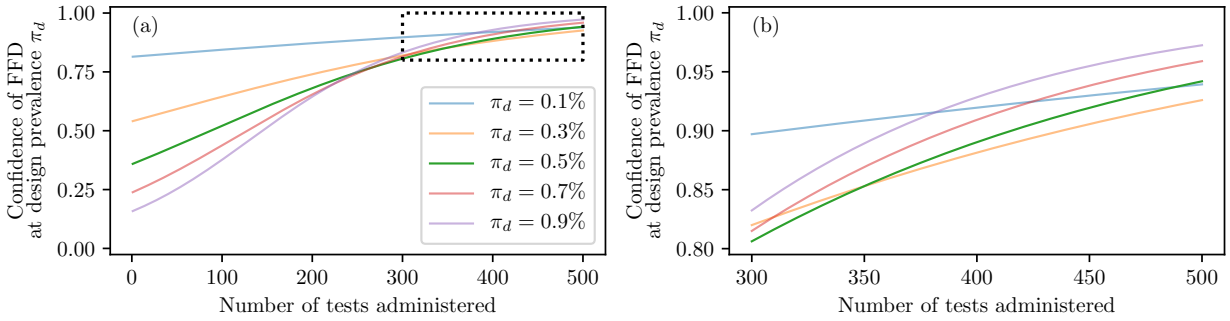


Figure 17: *Like Figure 6 except with sensitivity dropped from $Se = 0.7$ to $Se = 0.5$.*

Given the expectation that positive results will occur, due either to true or false positives, Section 4 extended the FFD approach to consider imperfect specificity and positive test results, aiming to develop an approach that instead considered the likelihood of an outbreak in the group of interest. Given the scenario of imperfect specificity, the outcome of the model provided a relatively marginal increase in assessing the likelihood of an outbreak as the number of positive tests increased (Figure 11b). Re-evaluating results with the lower sensitivity suggested in Section 5 only further decreases the ability to infer useful information from test results (Figure 18).

It is possible, despite the inability to assess the likelihood of point prevalence within the region of interest, that one may still be interested in trying to catch unknown cases with sentinel testing. This issue is examined in Section 5 and again, it is found that sentinel testing is likely of limited

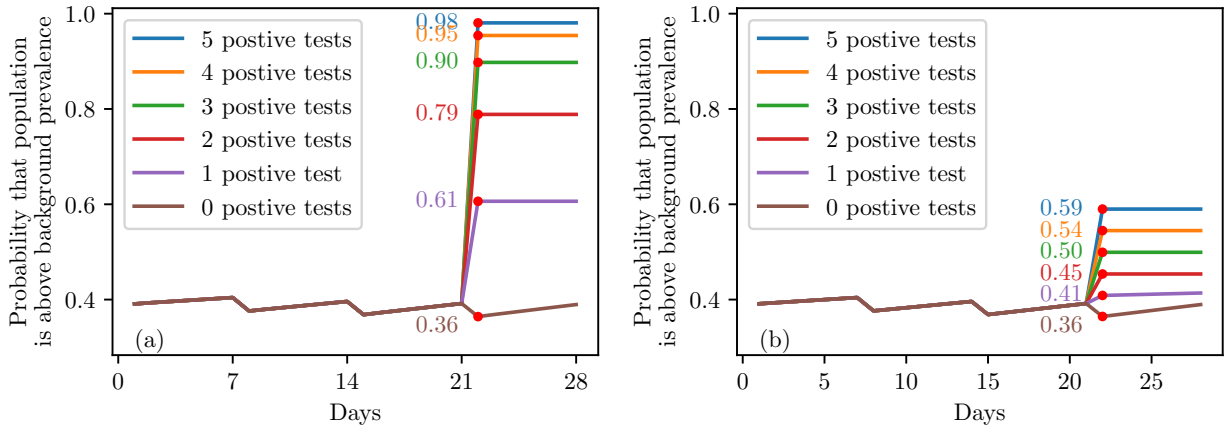


Figure 18: *Like Figure 11 except with sensitivity dropped from* $Se = 0.7$ *to* $Se = 0.5$.

utility. For the scenario where everyone is tested weekly, there is still only an estimated 44%–60% probability of catching an infection within the time window. (This result leads to the estimate of a 50% sensitivity in previous paragraphs of the discussion.) Looking at the scenario considered in this report for where 10% of the population is tested each week, the probability of catching an infection is an estimated 5%–7%. Based on the anticipated incidence rate, which suggests approximately one new case in the building each week, the sampling strategy would likely catch 3–4 of the roughly 50 cases each year. This low success rate is exacerbated when considering that there may also be false positive results. Assuming a specificity of $Sp = 0.99$, one could expect a false positive approximately once every two-to-three weeks. Under these assumptions, approximately 85%–89% of all positive tests are likely to be false positives.

An additional consideration is the interplay between test sensitivity and periodicity of testing. For example, as new rapid testing becomes more available, this may come at a cost of the heightened sensitivity that available from Polymerase Chain Reaction (PCR), often considered the "gold standard" for testing. The notion of frequent, low sensitivity testing has been proposed as the way ahead to contain the virus, specifically by controlling the number of non-symptomatic individuals [13]. With such tests, it may become possible to test large groups of individuals daily, but at the expense of sensitivity.

The methods described in Section 5 can be applied to consider that problem (Figure 19) to the trade-off between test sensitivity and test frequency if all individuals are tested. For Figure 19, the test window shown in Figure 13c remains steady but the sensitivity of the test is varied from 0 (no probability of detection) to 1 (perfect probability of detection). This figure is identical to Figure 15a, except that the label of the vertical axis is changed. This similarity is because in Figure 15a the uncertainty came from whether or not the infected person is tested, while in Figure 19 the uncertainty stems from whether the test detects the infection. An alternative approach, not considered in this report but aligned with the qualitative diagram in [13], would be have a time-dependent test sensitivity that relates to the viral load. Test sensitivity, particularly for a rapid test, is likely to have a greater—yet unknown—time dependence. Without having a good estimate of this time-dependence, however, it is necessary to view the results in Figure 19 qualitatively.

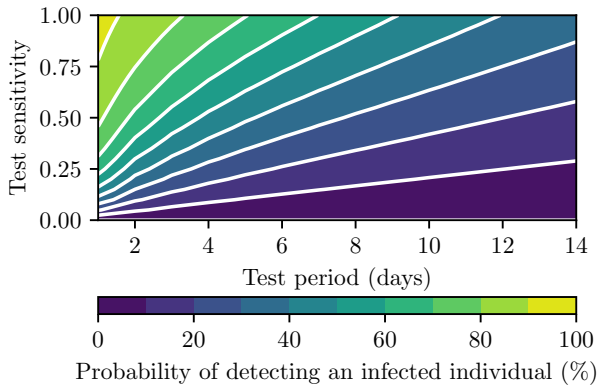


Figure 19: *Probability of detecting an infected individual as a function how often the individual is tested and sensitivity of the test administered.*

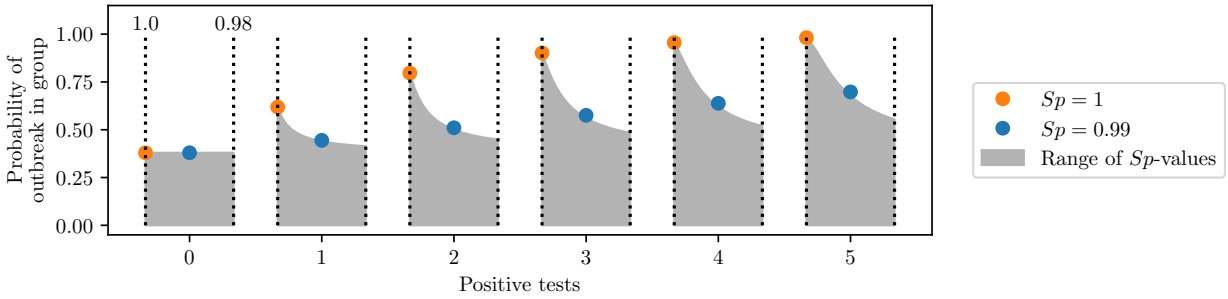


Figure 20: *Probability of an outbreak in a group with parameters like Figure 10, except for a broader range of Sp-values. Orange and blue dots show Sp-values for orange and blue bars in Figure 10.*

6.2 Effect of altering specificity

When considering imperfect specificity in this report, we have assumed an arbitrary value of $Sp = 0.99$. For the results in Sections 4 and 5, consideration of other potential values could alter findings.

Figure 10 and Table 3 show values for a single round of testing for $Se = 0.7$ and $Sp = 0.99$ or $Sp = 1$. Figure 20 shows results similar to Figure 10 except that the probability of being above background prevalence as a function of the number of positive tests shows continuous values ranging from $Sp = 1$ (left end of area over each number of tests) to $Sp = 0.98$ (right end of area). The leftmost values represent $Sp = 1$, are equivalent to the orange bars in Figure 10, and are depicted with orange dots. The middle values represent $Sp = 0.99$, are equal to the blue bars in Figure 10, and are depicted by blue dots. The rightmost values are equivalent to $Sp = 0.98$ and are lower than those considered earlier in this report. Results for a specificity of $Sp = 0.995$ are shown in Table 5.

Figure 21 shows how altering specificity could alter expectations for the expected number of false positive tests based on the scenario considered in this report (Panel a), and the positive predictive value of tests (Panel b). For example, if specificity were in fact $Sp = 0.995$, one would expect approximately 11 false positives per year (approximately one per month) and a 19%–24% chance that a positive test would reflect an infected individual. Nonetheless, the number of missed infections (93%–95%) would remain unchanged.

Table 5: *Reproduction of Table 3 but with an additional line for* $Sp = 0.995$.

Test	Number of positive tests					
specificity	0	1	2	3	4	5
$Sp = 0.99$	38%	44%	50%	56%	62%	68%
$Sp = 0.995$	38%	48%	57%	66%	74%	81%
$Sp = 1.00$	38%	62%	80%	90%	96%	98%

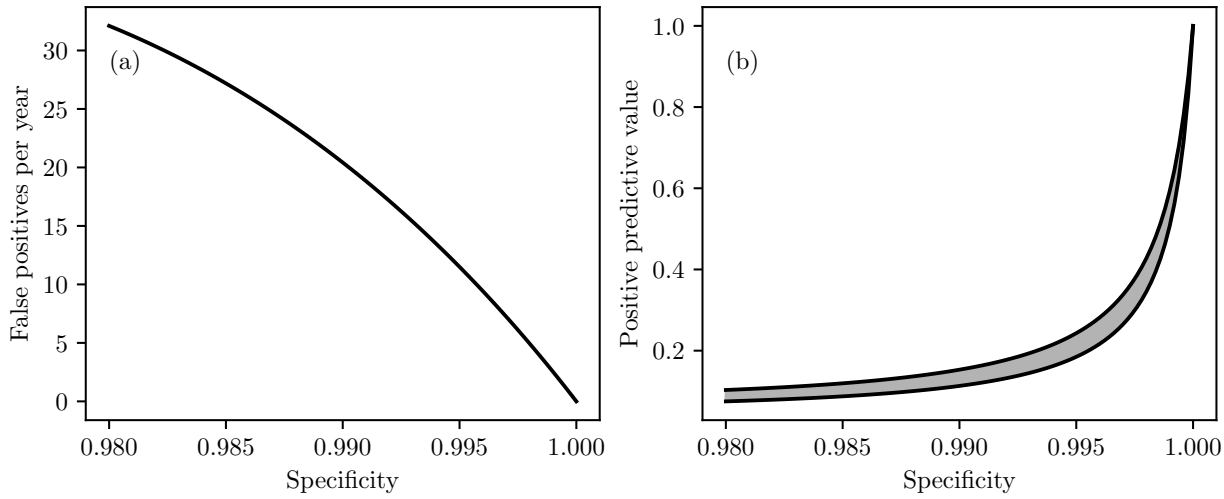


Figure 21: *Effect of specificity on (a) expected number of false positives per year; and (b) positive predictive value of tests. The region in Panel b reflects the bounds considered in Figures 14b and 14c.*

Increasing specificity does decrease the number of false positives, but they will likely continue to cloud results. In addition to any errors due to testing methodologies that could lead to false positives, asymptomatic individuals who are no longer infectious, for example, may still test positive. However, even in the case with no false positives, while there may be some possibility of catching an infection, it will still be challenging to use the approach to screen the population for cases, as demonstrated by the number of missed infections for a weekly testing strategy that tests 10% of the population.

While false positives may not be a major concern from a public health perspective, they are important from a building-management one. Presumably, individuals in the workplace at this time are there because they cannot complete their tasks from home. If there will be an ongoing need to quarantine and isolate not due to infected individuals but due to positive test results, then one must plan and manage their workforce accordingly to ensure continuity of operations.

6.3 Limitations and potential future work

In this report, we have attempted to review an existing surveillance method initially reported as a potential for tracking COVID-19 (Section 3), and to develop an extension to consider false positives (Section 4), before proposing that most sentinel testing may not be effective screening approach for an office building or other similarly sized group unless a large fraction of the population is tested on a very regular bases (Section 5). Nonetheless, there remain a number of areas where this work could be improved, including the following.

First, the method developed in Section 4 would likely benefit from a stronger recursive Bayesian filtering approach. The approach applies an admittedly simple approach, due primarily to the limited resources available when developed. Revisiting the problem with a more robust Bayesian approach would improve confidence in the results, and sentinel testing could potentially yield more information with the application of more sophisticated methods.

Second, the assumptions presented in Section 2 provide rough estimates that likely provide results at the appropriate order of magnitude, but could be refined. For example, the approach used to estimate prevalence by Horn [7] provides Bayesian estimates of a number of parameters. Using estimates from such a model could help to paint better uncertainty bounds on the results presented herein.

Last, a full simulation combining incidence within the population, different spreading scenarios, and application of a sentinel testing program could combine to provide another indicator of what can and cannot be identified through such methods. This approach would allow one to start to consider, e.g., variation in point prevalence over time and how this influenced variation in sentinel testing efficacy through time.

6.4 Concluding remarks

Ultimately, for the problem of a small office building examined in this report, it seems unlikely that sentinel testing will prove to be a worthwhile expenditure of resources given the assumed prevalence, unless it becomes possible to test the entire population almost daily. (Testing everyone weekly may still only catch 50% of infections, based on estimates derived in this report.) Instead, focusing on other mitigating efforts such as limiting meetings and minimizing other opportunities for contact are more likely to limit the probability of an outbreak within a building. Ensuring rapid and effective contact tracing (study underway) is also likely to mitigate COVID-19-related risks in the workplace more effectively than sentinel testing.

References

[1] Levesque, J., Maybury, D., and Shaw, D. (2020), A model of COVID-19 propagation based on a gamma subordinated negative binomial branching process, *Journal of Theoretical Biology*. Accepted November 2020.

[2] Doung-ngern, P., Suphanchaimat, R., Panjangampatthana, A., et al. (2020), Case-control study of use of personal protective measures and risk for SARS coronavirus 2 infection, Thailand, *Emerging Infectious Diseases*, 26, 2607–2616.

[3] Mirshak, R. and Pearce, T. (2020), Modelling the incidence of coronavirus disease 2019 (COVID-19) in Building 2 due to community spread, (Scientific Letter, DRDC-RDDC-2020-L229) Defence Research and Development Canada.

[4] Horn, S., Mirshak, R., Guillouzic, S., and Pelletier, E. (unpublished), COVID-19 mitigation for warships.

[5] World Health Organization, Sentinel Surveillance (online), https://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/sentinel/en/ (Access Date: 11 November 2020).

[6] Dekking, F., Kraaikamp, C., Lopuhaä, H., and Meester, L. (2005), A Modern Introduction to Probability and Statistics: Understanding Why and How, Springer Texts in Statistics, Springer.

[7] Horn, S. (2020), COVID-19 Point-Prevalence Map (online), https://stevenhorn.gitlab.io/covid-prevalence/map (Access Date: 30 September 2020).

[8] Foddai, A., Lubroth, J., and Ellis-Iversena, J. (2020), Base protocol for real time active random surveillance of coronavirus disease (COVID-19) - Adapting veterinary methodology to public health, *One Health*, Vol. 9. https://doi.org/10.1016/j.onehlt.2020.100129.

[9] Cameron, A. (2012), The consequences of risk-based surveillance: Developing output-based standards for surveillance to demonstrate freedom from disease, *Preventative Veterinary Medicine*, 105, 280–286.

[10] Cannon, R. (2002), Demonstrating disease freedom–combining confidence levels, *Preventative Veterinary Medicine*, 52, 227–249.

[11] Mirshak, R. and Horn, S. (2020), The probability of a non-symptomatic individual bringing a coronavirus disease 2019 (COVID-19) infection into a group: Methods and an example for the NORAD/USNORTHCOM Command Center (N2C2), (Scientific Letter, DRDC-RDDC-2020-L116) Defence Research and Development Canada.

[12] Guillouzic, S., Mirshak, R., and Sirjoosingh, A., Likelihood of undetected COVID-19 infection in a group: Effect of quarantining and testing (online), http://diaenterprisepublic.canadacentral.cloudapp.azure.com/COVID19MissedInfections.html (Access Date: 3 July 2020).

[13] Mina, M. J., Parker, R., and Larremore, D. B. (2020), Rethinking Covid-19 Test Sensitivity–A Strategy for Containment, *New England Journal of Medicine*. 10.1056/NEJMp2025631.

[14] Kobelia, M., Email with subject "Confidence level based testing summary," dated 6 May 2020.

[15] Schofield, S. and Lindsay, R., Private communication.

[16] Mirshak, R., Guillouzic, S., Horn, S., and Sirjoosingh, A. (unpublished), Estimating the probability of an unknown COVID-19 infection in groups with and without perfect sequestration.

[17] Cameron, A. and Baldock (1998), A new probability formula for surveys to substantiate freedom from disease, *Preventative Veterinary Medicine*, 34, 1–17.

[18] Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D., and Lessler, J. (2020), Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure, *Annals of Internal Medicine*. https://doi.org/10.7326/M20-1495.

[19] United States Center for Disease Control and United States Office of the Assistant Secretary for Preparedness and Response, Planning Parameters for COVID-19 Outbreak Scenarios (online), https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html (Access Date: 23 May 2020).

[20] Lauer, S. A., Grantz, K. H., Bi, Q., et al. (2020), The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application, *Annals of Internal Medicine*, 172, 577–582. https://doi.org/10.7326/M20-0504.

[21] Bullard, J., Dust, K., Funk, D., et al. (2020), Predicting infectious SARS-CoV-2 from diagnostic samples, *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciaa638.

[22] Cameron, A., Barford, K., Martin, T., Greiner, M., and Sergeant, E. (2003), Demonstrating freedom from disease using multiple complex data sources. Project Report DVI J.nr. 66032-0069.

[23] Sergeant, E. (2018), Epitools Epidemiological Calculators, Ausvet (online), https://epitools.ausvet.com.au/ (Access Date: 3 July 2020).

[24] Sergeant, E. (2018), Epitools - Sample size required to achieve target confidence of freedom, Ausvet (online), https://epitools.ausvet.com.au/freedomssthree (Access Date: 3 July 2020).

[25] Martin, P., A.R. Cameron, M., and Greiner (2007), Demonstrating freedom from disease using multiple complex data sources. 1. A new methodology based on scenario trees, *Preventative Veterinary Medicine*, 79, 71–97.

[26] Schofield, S., Private communication, 3 September 2020.

List of symbols/abbreviations/acronyms/initialisms

CDC United States Centers for Disease Control and Prevention

COVID-19 coronavirus disease 2019

FFD Freedom from Disease

NPI non-pharmaceutical intervention

NPV negative predictive value

PCR Polymerase Chain Reaction

PPV positive predictive value

SARS-Cov-2 severe acute respiratory syndrome coronavirus 2

WHO World Health Organization

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or a tasking agency, is entered in Section 8.)

DRDC – Centre for Operational Research and Analysis
Carling Campus, 60 Moodie Drive, building 7S.2, Kanata, ON K1A 0K2, Canada

2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.)

CAN UNCLASSIFIED

2b. CONTROLLED GOODS

NON-CONTROLLED GOODS
DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Challenges of sentinel testing for early detection of coronavirus disease 2019 (COVID-19) infections: Application to office buildings or other similar environments

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used. Use semi-colon as delimiter)

Mirshak, R.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total cited in document.)
December 2020	45	26

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter)

Scientific Report

8. SPONSORING CENTRE (The name and address of the department project or laboratory sponsoring the research and development.)

DRDC – Centre for Operational Research and Analysis
Carling Campus, 60 Moodie Drive, building 7S.2, Kanata, ON K1A 0K2, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable contract number under which the document was written.)
04k	

10a. DRDC DOCUMENT NUMBER	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2020-R127	

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

None

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

COVID-19; sentinel testing; statistical modelling; surveillance testing

13. ABSTRACT/RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)

This Scientific Report examines the utility of regular sentinel testing for monitoring COVID-19 activity in an office building or other similar setting where a group of individuals congregate but also mix with the general population. This is different from traditional sentinel testing where the objective is to understand the prevalence of a large region. Here, the goal is to ascertain the health of the building population relative to that regional scale.

Two methods are considered. The first is known as FFD, which was developed in veterinary medicine to examine livestock infections across a large region and was recently suggested as a tool to demonstrate absence of COVID-19 on a military base. The second is an extension that accepts the virus may be within the group and, considering the effect of false positives, provides insights on the point prevalence in the group relative to the background point prevalence of the community. Both methods apply a Bayesian approach to make these inferences.

Due to the mixing of the group of interest with the general population, the probability of introduction of the virus between rounds of testing is quite high. As a result, unless testing is at extremely high levels, it is not possible to get a high confidence that the point prevalence in the group is low. As a result, neither method suggests sentinel testing will provide much protection for or information about the group.

Additional investigation shows that the time-dependent aspect of testing, i.e., that an individual must be tested in a timely manner for sentinel testing to be effective, further exacerbates the problem. Ultimately, true positive tests from sentinel testing are likely to be overwhelmed both by false positives and by individuals who are identified clinically after symptoms present.

Although exploring other options is beyond the scope if this work, mitigation measures that limit spreading opportunities before an infection is known and aggressive contact tracing after an infection is identified are likely to be more effective at protecting the group from an outbreak than a limited sentinel sampling strategy.

Ce rapport scientifique examine l'utilité des tests sentinelles réguliers pour surveiller l'activité de la maladie à coronavirus 2019 (COVID-19) dans un immeuble de bureaux ou autre lieu similaire où les personnes se regroupent, mais se mélangent aussi avec la population générale. Cela diffère des tests sentinelles traditionnels dont l'objectif est de comprendre la prévalence de la maladie dans une grande région. Dans le présent document, l'objectif est de vérifier la santé de la population de l'immeuble par rapport à celle de la population à l'échelle régionale.

Deux méthodes sont envisagées. La première méthode est connue sous le nom de validation de l'absence de la maladie, qui a été élaborée en médecine vétérinaire pour examiner les infections du bétail dans une grande région et a récemment été proposée comme outil pour démontrer l'absence de COVID-19 dans une base militaire. La deuxième méthode est une extension qui accepte que le virus puisse se trouver dans le groupe et, compte tenu de l'effet des faux résultats positifs, donne des indications sur la prévalence instantanée dans le groupe par rapport à la prévalence instantanée de base de la communauté. Les deux méthodes appliquent une approche bayésienne pour faire ces déductions.

En raison du mélange du groupe d'intérêt avec la population générale, la probabilité d'introduction du virus entre les séries de tests est assez élevée. Par conséquent, à moins que le nombre de tests effectués soit extrêmement élevé, il n'est pas possible d'obtenir une grande certitude que la prévalence instantanée dans le groupe est faible. Par conséquent, aucune des deux méthodes ne suggère que les tests sentinelles apporteront une grande protection du groupe ou des renseignements sur le groupe.

Une enquête supplémentaire montre que l'aspect temporel des tests, c'est-à-dire qu'une personne doit être testée en temps utile pour que les tests sentinelles soient efficaces, aggrave encore le problème. En définitive, les vrais résultats positifs des tests sentinelles risquent d'être dépassés à la fois par les faux résultats positifs et par les personnes qui sont testées cliniquement après la présence des symptômes.

Bien que l'exploration d'autres options dépasse le cadre de ces travaux, les mesures d'atténuation, qui limitent les possibilités de propagation avant qu'une infection ne soit connue et la recherche agressive des contacts après l'identification d'une infection, seront probablement plus efficaces pour protéger le groupe contre une éclosion qu'une stratégie d'échantillonnage sentinelle limité.