PLOS ONE

RESEARCH ARTICLE

On an optimal testing strategy for workplace settings operating during the COVID-19 pandemic

X. Hernandez[1]*, **S. Valentinotti**[2]

1 Instituto de Astronomia, Universidad Nacional Autónoma de México, CDMX, México, **2** Laboratorios Liomont S.A. de C.V., Adolfo López Mateos 68, CDMX, México

* xavier@astro.unam.mx

Abstract

High quality daily testing for the presence of the SARS-CoV-2 in workplace settings has become part of the standard and mandatory protection measures implemented widely in response to the current pandemic. Such tests are often limited to a small fraction of the attending personnel due to cost considerations, limited availability and processing capabilities and the often cumbersome requirements of the test itself. A maximally efficient use of such an important and frequently scarce resource is clearly required. We here present an optimal testing strategy which minimises the presence of pre-symptomatic and asymptomatic infected members of the population in a workplace setting, derived under a series of simplifying statistical assumptions. These assumptions however, retain many of the generalities of the problem and yield robust results, as verified through a number of numerical simulations. We show that reduction in overall infected-person-days, IPD, by significant percentages can be achieved, for fixed numbers of tests per day of 5% and 10% of the population, of 30% and 50% in the IPD numbers, respectively.

Citation: Hernandez X, Valentinotti S (2022) On an optimal testing strategy for workplace settings operating during the COVID-19 pandemic. PLoS ONE 17(3): e0264060. https://doi.org/10.1371/journal.pone.0264060

Editor: Cecilia Ximenez, Faculty of Medicine, National University of Mexico (UNAM), MEXICO

Received: March 11, 2021

Accepted: February 2, 2022

Published: March 2, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Within the context of the present COVID-19 pandemic, it has become clear that thus far, the most efficient strategy towards reducing the spread of the disease includes strict social distancing rules, reinforcing basic hygiene measures and the imposition of lockdown policies on the part of local and national governments. This last however, must clearly be tempered by the obvious need to keep essential workplace facilities operating. Examples of the above include hospitals, energy production facilities, food production and distribution infrastructure, and pharmaceutical industries, to mention but a handful of the most obvious such cases.

Continual operation of such facilities has firstly included the adoption of safety and hygiene protocols during working days, and crucially, strict Sanitary Checkpoints, protocols for the daily entrance of persons attending. Sanitary Checkpoints (SC) serve the purpose of identifying symptomatic individuals which can then be tested directly for the virus, or in any case, sent home for a certain safety period.

Unfortunately, in the present pandemic, SC implementation is a measure of only limited efficacy, due to the important contribution of pre-symptomatic and asymptomatic infected persons [1, 2]. It is now clear that a substantial fraction of transmissions are in fact the result of interactions with pre-symptomatic and asymptomatic individuals [3]. This has led to the implementation of testing strategies, where a certain sample of persons attending are selected for direct, accurate testing, typically through PCR or more recently antigen testing e.g. [4]. These accurate tests however, were initially not widely available and remain relatively scarce in certain areas, particularly outside big cities and in developing countries [5]. Further, such tests are somewhat cumbersome to implement, as well as expensive and generally not amiable to massive, daily implementation. Indeed, the need for such testing has given rise to a number of innovative and imaginative solutions, e.g. the use of mobile testing clinics as described in [6]. For the above reasons, in most facilities where PCR testing is used, only a small fraction of the regular work force is sampled on any particular day e.g. [7]. Thus, we have a situation where it is of the utmost urgency to apply optimal strategies to select the daily test sample, a scarce and expensive resource that must clearly be used to maximal efficiency.

It is important to note that the identification of asymptomatic persons is not only of relevance for the safe operation of an essential production facility, but also as a clear means of reducing the overall extent of the pandemic. As more infected individuals are identified anywhere (and hence identified for surveillance, contact tracing, early treatment and/or quarantine protocols), the more successful societies will be in controlling epidemics.

While the importance of small screening intervals has been pointed out [8], such approaches assume a copious availability of daily tests, a situation which is not always practicable. The optimal arrangement and practical implementation of SC facilities such that a large number of daily tests can be performed with a minimal disturbance and a maximal efficiency is an important topic which has been treated by various authors e.g. [9].

In this paper we present an optimal sample selection strategy which maximises the number of asymptomatic infected persons identified from a fixed population, under the restriction of a given fixed number of daily tests available. In the following section we present simple probabilistic arguments showing that a randomly selected daily sample, from which persons which have already tested negative within the immediate τ_I day period have been excluded, should minimise the total number of infected-person-days (henceforth IPD) over any fixed period of time. τ_I is the timescale over which the general population of infected persons is replaced by a new one, under the common assumption of a constant 15 day infection period, $\tau_I = 15$ days.

The value of probabilistic simulations not only in the modelling of the evolution of epidemics, but also for estimating the effects of interventions aimed at mitigating aspects of such events is well established. We here cite only a few recent examples in the context of the current pandemic, where the effectiveness of practical strategies is gauged through numerical modelling within particular imposed restrictions e.g. [10–13].

In our results section we implement a number of numerical simulations following idealised populations subject to a number of infection probabilities, percentages and evolution of pre-symptomatic and asymptomatic fractions and fraction of available daily tests. We show that indeed the optimal strategy presented results in a minimal number of IPD over any fixed period of time, and that this result is general to the epidemiological evolution, a constant or rising overall infection fraction in the population. The discussion addresses a number of caveats of the approach presented, as well as generalisations and expected developments.

While a number of idealised assumptions remain in the model, it can be shown that the conclusion is sufficiently robust to warrant attention as a further element in aiding the control of the present pandemic.

Probabilistic developments

We begin with simple considerations, imagine a constant infection fraction phase of the epidemic at a particular location, where the average fraction of infected individuals, I_f, is a small constant number. If the infection period for all, symptomatic or asymptomatic persons, is a constant τ_I number of days, and if we imagine a cohort of simultaneously infected persons, it is clear that over the period over which they are infected, they must each infect one healthy individual, to ensure I_f remains a constant. Hence, assuming the probability of becoming infected in any single day for a healthy person in the general population is a constant, P_I, and taking a linear approximation for the cumulative infection probability, the probability of becoming infected after a τ_I day period, P_τ, will satisfy:

$$P(\tau) = \tau_I P_I = I_f, \tag{1}$$

such that after a τ_I day period, each healthy individual has a chance I_f of having been infected, and therefore, the next cohort of infected individuals again represents the same fixed I_f fraction of the total. If the fraction of infected individuals is a constant I_f, it is reasonable to assume that on average, the probability that a healthy individual has of becoming infected on any particular day, will be a constant, P_I. From the above linear approximation, given values of I_f and τ_I, the daily infection probability can be estimated as:

$$P_I = I_f/\tau_I \tag{2}$$

It is clear that cohorts of infected individuals will not be temporally exclusive, but will occur in a scrambled fashion over time, not altering the above equation, provided I_f and τ_I remain constant over time. If additionally, a certain person had a negative test result on a particular day, we shall assume that the probability of that individual being infected that day is zero, i.e. no false negatives are assumed on tests being performed. Thus, the probability of that individual becoming infected the following day becomes P_I, and the probability of the individual not becoming infected on the day following his negative test is $(1 - P_I)$. The probability of not becoming infected after two days following the negative test is now $(1 - P_I)^2$, and in general, on average, after n, days, the probability of being infected becomes a non-linear function given by:

$$P(n) = 1 - (1 - P_I)^n, \tag{3}$$

which corrects the linear approximation given previously. Again, over a τ_I day period, the above probability must yield I_f, if this fraction is to remain constant, so that:

$$1 - (1 - P_I)^{\tau_I} = I_f \tag{4}$$

From this equation P_I can be calculated, for given values of I_f, and τ_I. Notice that for $P_I \ll 1$, a Taylor expansion of Eq 4 to first order recovers the zero-order intuitive result of Eq 2, as is evident from writing the term in brackets in Eq 4 as $(1 - P_I)^{\tau_I} = 1 - \tau_I P_I + \mathcal{O}(P_I^2)$, yielding $\tau_I P_I = I_f$, to first order in P_I, which corresponds also to Eq 1. In order to develop an optimal testing strategy, we shall assume that we have a fixed number of persons attending the facility in question every day, having, in the absence of any test, an equal infection probability as the general population, I_f. Thus, the first order testing strategy, if N_T is the number of tests available every day, is simply to draw a random sample of N_T members from the attending population at the workplace setting every day. However, notice that if someone tested negative on one particular day, the probability that this person is infected will begin to grow from zero on the days following the negative test, while the probability that a random member of the population is

infected, prior to any tests, will be I_f. Thus, we should endeavour to test attending persons having the highest probability of being infected, so that the limited number of tests available are used optimally towards identifying more effectively the infected members. Clearly, a member who has tested negative on a particular day should be excluded from the testing pool until the probability of this member being infected becomes again equal to that of the average population. Such a member should be excluded from the testing pool for a period n_{ex} such that:

$$1 - (1 - P_I)^{n_{ex}} = I_f. \tag{5}$$

Clearly, Eqs 4 and 5 are identical, and therefore,

$$n_{ex} = \tau_I, \tag{6}$$

which is the main result of this section. A graphic representation of this is presented in Fig 1, which shows the logarithm of the probability of being infected after an n day period, $P(n)$ of Eq (3), for $P_I = 6.69798 \times 10^{-4}$. This value of P_I corresponds to a steady infection fraction of $I_f = 0.01$ and $\tau_I = 15$, as resulting from Eq 4 written as $P_I = 1 - (1 - I_f)^{1/\tau_I}$. Under the $P_I \ll 1$ approximation of Eq 2 we obtain $P_I = 0.01/15 = 6.666687 \times 10^{-4}$. Although Eq 2 is clearly a good approximation to the value given by Eq 4 for the small values of P_I expected, we use throughout exact values from Eq 4.

As can be seen from the figure, the probability of being infected grows from zero on the day of the negative test, gradually with n, and converges to a value of 1 as n tends to infinity. The horizontal line shows the value of $I_f = 0.01$, the average probability of being infected for the general population. Clearly, $P(n)$ overtakes I_f precisely at $n = \tau_I = 15$ days. At the start of 2021, the 15 most affected countries reported close to 1/1000 COVID-19 deaths per capita [14], which assuming a 0.5% infection fatality rate yields 20% as the fraction of the population which has been infected in these countries, on average. Assuming further a 10-month duration for the worst of the pandemic and a 15-day infection period, yields an average infected fraction of 1% for these countries. We hence take 1% for this variable, as a broad average reference value. Although approximately constant values of I_f over time have been observed for the current pandemic on a variety of places and times, e.g. the slowly rising bursts seen in some countries, or the extended troughs between waves often seen in others [14], in general, I_f will be a function of time. While to the accuracy presently available τ_I appears to be close to a constant, a situation where $I_f = I_f(t)$ will be more common, leading also to $P_I = P_I(t)$. Interestingly, in such a case, the factors $(1 - P_I)^x$ in Eqs 4 and 5 will both be replaced by the same function of time, namely:

$$(1 - P_I)^x \rightarrow \prod_{i=i_0}^{i=x} (1 - P_I(i)) \tag{7}$$

where i_0 is a relevant initial time index. Clearly, again, the optimal testing strategy remains unchanged, with $n_{ex} = \tau_I$ This conclusion applies whenever P_I is a slowly varying function of time, in comparison to τ_I. Any sudden spikes, e.g. the influx of a considerable fraction of infected persons added to the total population, will clearly invalidate the argument above. It is of course possible that a member testing negative on a particular day becomes infected on the following one. If such case happens to be one of the asymptomatic ones, he will be allowed to attend while infected for the 15-day period over which he is infected, and presumably infectious. This however is less likely than that an average member who has not been tested might be infected, as can be seen from Fig 1. With a limited number of tests per day, one can at best

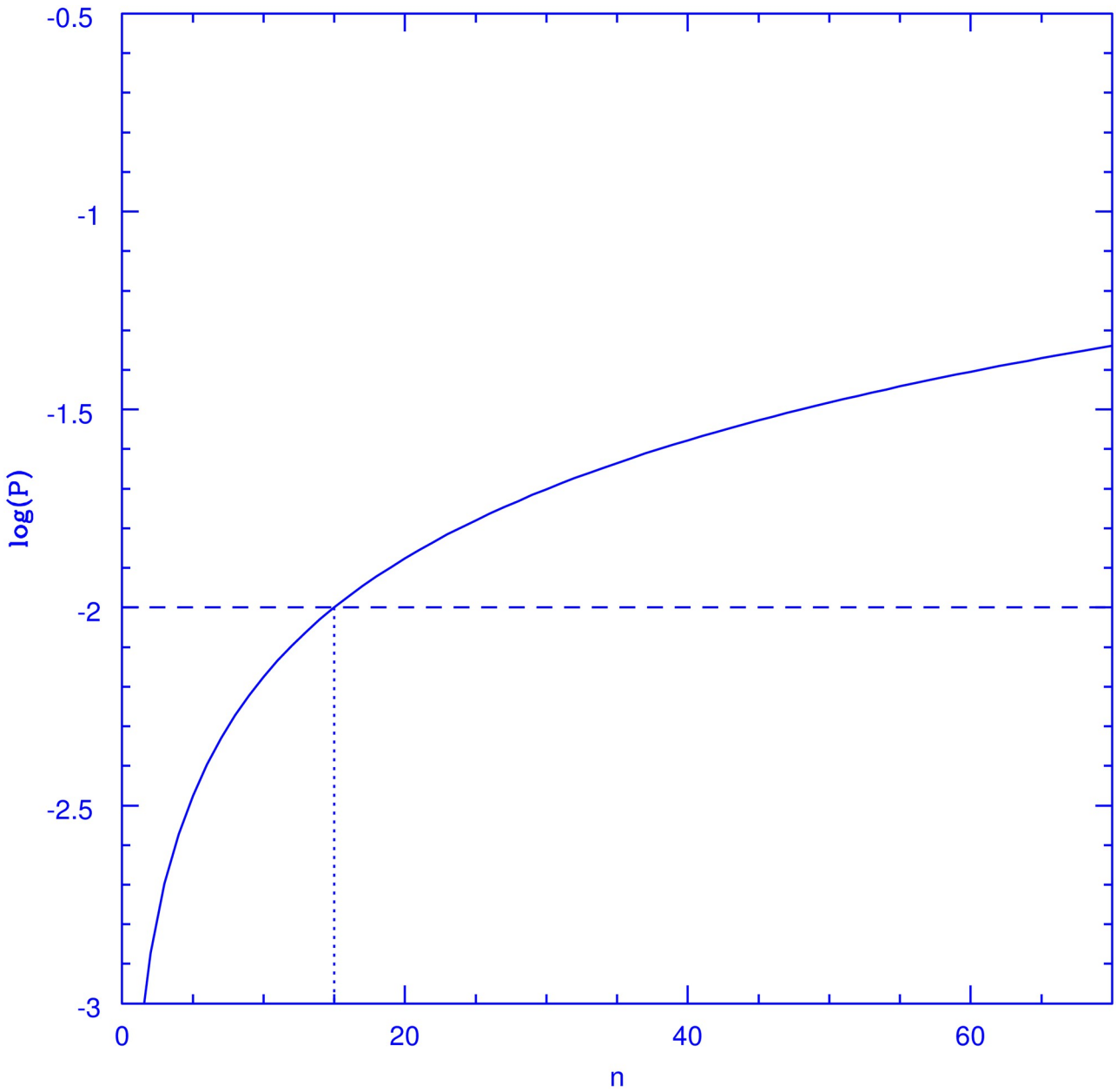

Fig 1. Probability of being infected *n* days after a negative test. The figure shows the base 10 logarithm of the probability for an individual which tested negative on day zero of being infected on subsequent days, $log_{10}\, P(n)$ where $P(n) = 1 - (1 - P_I)^n$, for $P_I = 6.69798 \times 10^{-4}$, corresponding to a steady infection fraction of I_f 0.01 and $\tau_I = 15$. This probability remains below that for the average population, 0.01, for all days before $n = 15$, after which, it remains above this number indefinitely.

https://doi.org/10.1371/journal.pone.0264060.g001

minimise the number of infected persons attending, but driving this number to zero is unfortunately impossible, unless the daily testing of everyone might become an option.

Finally, we consider the effects of having a finite total sample. It is clear that as time goes on, for certain fractions of the total population being tested, it could well happen that on a certain day there are not enough members who have not been tested within the previous 15 days to complete the N_T tests available that day. Given the positive definitive character of the probability function shown in Fig 1, for any value of its parameters, it follows that the optimal

Table 1. Reference table.

Parameter	Definition
I_f	Infection fraction in the general population
P_I	Daily average probability of infection in the general population
τ_I	Duration of the infection period
N_T	Number of tests performed every day
n_{ex}	Number of days after a negative test over which an individual is ideally excluded from testing sample
IPD	Number of Infected-Person-Days recorded over a 100 day period

Definition of parameters of the probabilistic model and numerical simulations.

https://doi.org/10.1371/journal.pone.0264060.t001

sampling strategy becomes to exclude from the daily testing pool any members who have been tested within the previous 15 days, while if that rule cannot be met without reducing N_T, the sample should be augmented by adding all members who have gone 14 days without testing. If this again does not allow to complete the N_T tests available, all those who have gone 13 days without testing are considered, and so on.

Notice that once finite population effects begin to appear and the available tests have to be made up by including members not having been tested over periods smaller than τ_I, the testing sample becomes drawn from a mixed population, including members who have not been tested over a range of recent days. Hence, taking n_{ex} values in this range will result in a softening of the resulting IPD(n_{ex}) curves, the expected infinite population minimum at $n_{ex} = \tau_I$ can become broad, with a complex dependence on the details. The specifics of this will depend on the fraction of tests available and the epidemiological details, as those will determine the average rate at which infected persons are found and sent home, which in turn affects the numbers remaining, from which the daily sample is to be drawn. We end this section with a small reference table summarising all of the parameters of the probabilistic model and simulations used, Table 1.

Results

We now describe and show the results of a number of numerical simulations following sample populations under a variety of assumptions, where we can assess the generality of the scheme presented. We model a population of 1,000 members evolving under the following rules: Initially all members are assigned a healthy or infected status with a probability $(1 - I_f)$ or I_f, respectively, with $I_f = 0.01$. Of those assigned as infected, a random sample of 25% are assigned as displaying symptoms and the rest as not displaying any symptoms, these last are a mix of the pre-symptomatic and the asymptomatic ones. During every following day, the epidemiological evolution considers a probability of becoming infected of $P_I = 6.69798 \times 10^{-4}$ for each healthy member. Again, 25% of those newly infected are assigned as displaying symptoms at the onset. Also, all those infected and not displaying symptoms are assigned, only on their 5th infected day, a 60% probability of passing from infected and not displaying symptoms to infected and displaying symptoms. This fixes a final asymptomatic fraction of $0.4 \times 0.75 = 0.3$, in accordance with recent estimates [2]. Further, any infected members having spent more than 14 days in this state are returned to a healthy status. This ensures a steady $I_f = 0.01$ value, on average.

Then the intervention is imposed, any members displaying symptoms are sent home for a 14-day period, under the assumption of an efficient SC. Then, a testing sample as described in the previous section is constructed, which will have a variable number of N_S members from which a fixed N_T members are selected for testing at random every day. Fig 2 shows

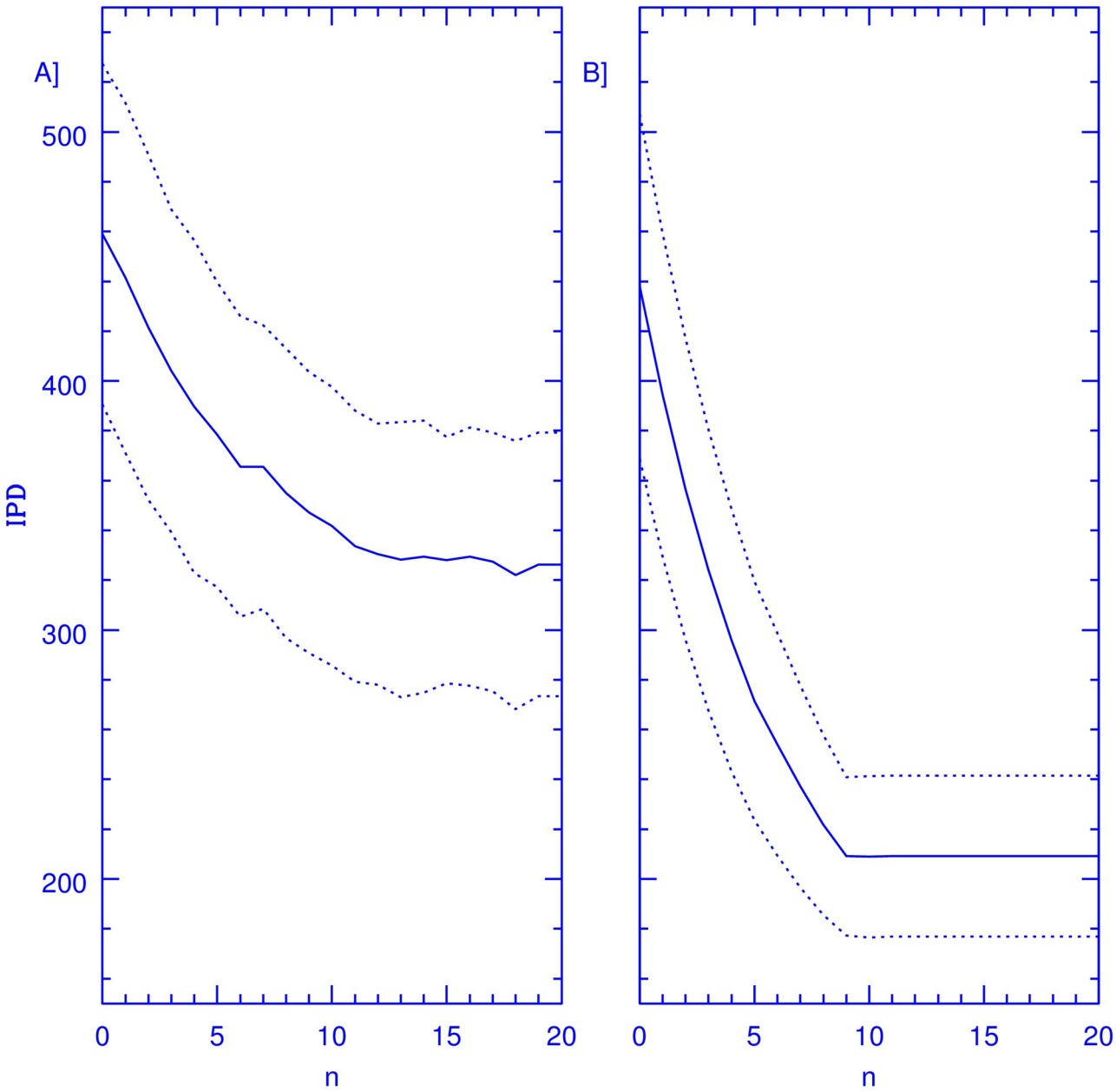

Fig 2. Resulting IPD comparisons as a function of n_{ex}, constant I_f. The figure shows IPD values over a 100-day period, solid curves, obtained for a population of 1000 members attending a facility where 50 (panel A) and 100 (panel B) PCR tests are performed every day on a random sample of members, from which persons which test negative on a particular day are excluded for the subsequent n_{ex} days, as a function of n_{ex}. Members testing positive are sent home for a 14 day period. The epidemiological model corresponds to a value of $I_f = 0.01$ which remains constant over time. The dotted curves show the stochastic variance inherent to the problem, giving 1σ intervals over 10,000 realisations, this last corresponding to the dotted lines. An efficient SC is also assumed.

https://doi.org/10.1371/journal.pone.0264060.g002

results for $N_T = 50$ and $N_T = 100$, left and right panels respectively. This is done for a fixed n_{ex} value, and the simulation is run for 200 days. Every day the number of infected persons attending is recorded to determine the IPD resulting. Given the stochastic nature of the problem, the whole 200-day simulation is repeated 10,000 times to gauge the intrinsic

variance present. This is then repeated for various values of n_{ex} in the range shown in Fig 2. The solid lines give the average over the 10,000 realisations of the IPD values obtained over the second 100-day period. The first 100-day period is run to allow for a steady state to develop. The dotted lines give the 1σ intervals for the IPD values shown, over the 10,000 realisations, which for the parameters modelled are sufficient to reach convergence, further increase in the number of realisations considered yields only marginal differences from the results reported.

In absence of any intervention, one expects an average value of IPD = $1{,}000 \times I_f \times 100 = 1{,}000$ indeed, we get 1000±118 for this quantity over the 10,000 realisations. The introduction of the SC alone reduces these values to 484±73, a reduction of a little more than a factor of two. If one then also includes a random sampling of 50 tests from which none of the members attending are excluded, we obtain an IPD = 458±68, as shown by the $n_{ex} = 0$ point in the left panel of Fig 2. We see that the pure random sample achieves only a small further reduction in IPD numbers, as persons recently tested (and hence having a very small probability of being infected) have an equal chance of being selected for testing as persons not having been tested over a longer period. As the number of days after test which result in exclusion from testing is increased, a clear drop in IPD appears. Then, optimising further by the careful construction of the testing sample described, reduces the IPD values still further, down to 322±54 in the broad minimum around $n_{ex} = 15$. Hence, a very significant reduction of 30% in IPD numbers, at constant number of daily tests performed, is achieved merely by excluding recently tested persons from the daily samples.

In the right panel of Fig 2 we see that as the testing sample grows to $N_T = 100$, the optimisation through the exclusion of recently tested members yields much more important results, with the drop from $n_{ex} = 0$ to $n_{ex} = 15$ now being of a very sizeable 52% of the IPD numbers found at $n_{ex} = 0$, showing the power of the sampling strategy presented. In this case however, we see clearly the finite sample effects appearing in the convergence at $n_{ex} = 9$. Beyond this point (for the 10% of the total population being tested daily) it becomes impossible to find a daily sample not including persons having not been tested over a $n > 9$ day period, and the method saturates. Notice that for $N_T = 50$ and $n_{ex} > 10$ the total IPD numbers become even smaller than what is obtained for the double number of daily tests but $n_{ex} = 0$.

Finally, Fig 3 is equivalent to Fig 2, but gives results for simulations where the underlying epidemiological model is one where P_I rises over time with a constant doubling timescale of 50 days, being equal to that of the case summarised in Fig 2 at 100 days. In this case, in the absence of any intervention we obtain IPD = 2077±154. The introduction of an efficient SC reduces these numbers to IPD = 1329±115. Again, the reduction achieved by the optimisation of the testing sample is important in both cases, when $N_T = 50$, of 28.4% of the $n_{ex} = 0$ values. Yet, in going to $N_T = 100$, shown in the right panel of Fig 3, the corresponding reduction is now of a much more significant 51%, achieved without increasing the number of tests performed, merely through an optimised testing strategy. Again, saturation through finite sample effects becoming apparent at $n_{ex} = 9$. As in the previous case, the convergence of the $N_T = 50$ simulation for large n_{ex} values actually occurs at IPD values smaller than what is obtained in the NT = 100 one at $n_{ex} = 0$. Also illustrative of the strength of the method presented is that the fractional reduction in total IPD numbers shown, is in all cases stronger in going from $n_{ex} = 0$ to optimal values, than what results from going from the pure SC case to the $N_T = 50$ and $N_T = 100$ ones at $n_{ex} = 0$.

All simulation input parameters, and resulting IPD numbers are summarised in Table 2. IPD_0 gives the infected-person-days in the absence of any SC or testing strategy, IPD_{SC} gives the corresponding value in the presence of the SC alone, IPD_{Ra} values when the SC is present, and N_T tests are performed at random from the entire population, and finally, IPD_{Op}, the

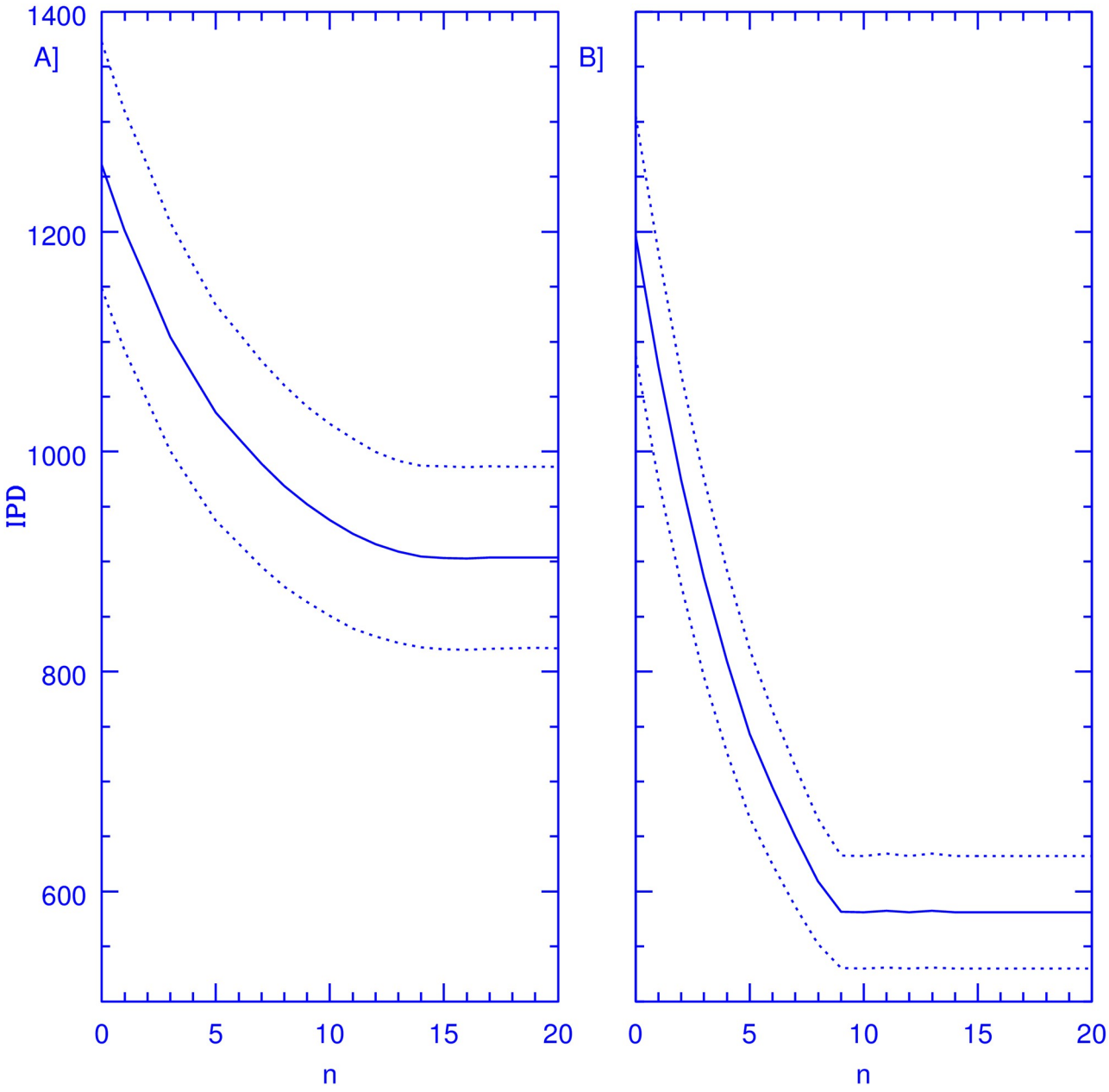

Fig 3. Resulting IPD comparisons as a function of n_{ex}, raising P_I. The figure shows IPD values over a 100-day period, solid curves, obtained for a population of 1000 members attending a facility where 50 (panel A) and 100 (panel B) PCR tests are performed every day on a random sample of members, from which persons which test negative on a particular day are excluded for the subsequent n_{ex} days, as a function of n_{ex}. Members testing positive are sent home for a 14-day period. The epidemiological model corresponds to a value of $P_I(t)$ which grows linearly having a doubling timescale of 50 days and starting at $P_I = 6.69798 \times 10^{-4}$. The dotted curves show the stochastic variance inherent to the problem, giving 1σ intervals over 10,000 realisations, this last shown by dotted lines. An efficient SC is also assumed.

https://doi.org/10.1371/journal.pone.0264060.g003

results when the SC is present, and N_T tests are performed at random after removing from the testable population individuals which tested negative within a previous 15 day interval, or the largest number of days available below 15 if the previous condition can not be met. Lastly, Sk and Ku give the third and fourth moments of the IPD distributions at convergence, calculated

Table 2. Simulation input parameters and results.

Case	P_I	N_T	IPD_0	IPD_{SC}	IPD_{Ra}	IPD_{Op}	Sk	Ku
Fig 2A	Fixed	50	1000 ± 118	484 ± 73	458 ± 68	322 ± 54	0.19	3.08
Fig 2B	Fixed	100	1000 ± 118	484 ± 73	437 ± 69	209 ± 32	0.15	3.02
Fig 3A	Rising	50	2077 ± 154	1329 ± 115	1261 ± 112	903 ± 83	0.12	3.07
Fig 3B	Rising	100	2077 ± 154	1329 ± 115	1196 ± 109	581 ± 51	0.09	2.98

Simulation input parameters and resulting IPD numbers. IPD_0 gives the infected-person-days in the absence of any SC or testing strategy, IPD_{SC} gives the corresponding value in the presence of the SC alone, IPD_{Ra} values when the SC is present, and N_T tests are performed at random from the entire population, and finally, IPD_{Op} the results when the SC is present, and N_T tests are performed at random after removing from the testable population individuals which tested negative within a previous 15 day interval, or the largest number of days available below 15 if the previous condition can not be met.

https://doi.org/10.1371/journal.pone.0264060.t002

through:

$$Sk = \frac{E[(x-\mu)^3]}{(E[(x-\mu)^2])^{3/2}}, Ku = \frac{E[(x-\mu)^4]}{(E[(x-\mu)^2])^2}, \quad (8)$$

where μ gives the mean of a distribution of variable x and $E[f(x)]$ the expectation value of any function $f(x)$.

A fairly general feature is that the reduction in IPD numbers obtained becomes fractionally smaller as n_{ex} increases, this is a direct consequence of the second derivative of the probability function of Eq 5 being negative, as shown in Fig 1. The convergence of this probability function as n_{ex} grows means that the largest drops in IPD numbers will come from the initial increases in n_{ex} away from zero, although the broad minimum remains at $n_{ex} = 15$, this last, provided no finite sample convergence appears.

The examples shown above are arbitrary and clearly extremely over simplified, not intended to represent any particular real situation. They do however serve to illustrate the usefulness of the scheme presented, as well as its generality and robustness to changes in the underlying epidemiological model. Notice from Table 2 the small effect which a 5% or 10% rate of daily testing has, if the testing sample is simply a fully random subset of the total population. IPD_{Ra} values gain very little on IPD_{SC} ones, it is in going from IPD_{Ra} to IPD_{Op} that the test performed yield an important effect. The actual reduction in IPD numbers will be sensitive to the details of the particular problem, but under a wide range of parameters, the conclusion of a daily random sample from which persons that have tested negative within the previous τ_I days were removed, as representing the optimal solution, will remain.

Assessing whether or not the effects of the optimal strategy presented are meaningfully beyond statistical variance, requires an appraisal of the shape of the resulting IPD distributions. We have also calculated the third and fourth moments of the IPD distributions throughout. The Skewness parameter is consistently very low, ranging from between 0.05 to 0.22 for all the evolution of all the simulations presented, i.e., far from the standard value of 1 which characterises important deviations from symmetry. The Kurtosis parameter was consistently within 0.1 of 3, hence, excess Kurtosis values within 0.1 of Gaussian. Hence, for example in the case of Fig 2A, we can be confident that a 2 sigma difference in IPD numbers implies that 98 out of 100 times the application of the optimal strategy will result in a reduction of total IPD numbers when using the optimal strategy in comparison with a pure random sampling. In the other cases the result is typically more significant, with differences of up to 5 sigma appearing between pure random and optimal testing. Final convergence ($n_{ex} = 15$) values for Skewness and Kurtosis for the four simulations described are given in Table 2.

Further, we have included the assumption of 100% sensitivity in the PCR tests, while in reality this appears to grow rapidly from zero with time since infection to reach a maximum at around 8 days followed by an initially slow drop with time e.g. see the appendix in [8] and references therein. Therefore, the conclusion presented will not be modified, as re-testing within the first few days will not only be unlikely to detect an infected person (their chances of being infected being still small) but also, as these tests will be largely "wasted" through their sensitivity being low. The effect of the steeper decline in test sensitivity after about 13 days will be negligible in cases where the daily test percentage of the population is larger than 10%, where the saturation mentioned above appears before this period. We have taken an asymptomatic fraction towards the lower range of inferred values [2], with the intention of showing clearly the potential of the method presented, under less than optimal conditions for it; clearly, the effect of the method will tend to zero as the asymptomatic fraction goes to zero, and increase substantially as this fraction increases.

The scalings resulting from changes in the total population, as expected from basic probabilistic considerations, are for average resulting IPD values which remain constant as a fraction of the total population considered, with the corresponding 1σ intervals scaling with the square root of the total population considered, all other parameters being equal.

Fatalities are not explicitly included in the model, but will not have any significant effects provided the total fatality rate is low and/or any fatalities are promptly replaced from the general population. Finally, we note that the method presented can obviously be used in conjunction with pooling strategies, which, whenever sample taking logistics and infection prevalence allow (see for example [15, 16]), permit increases in the total number of daily tests.

Discussion

We have presented a local epidemiological model where the average infection rate is assumed as determined by the overall infection rate of a global population of which the local model represents a fair sample. The intervention proposed does not aim at altering the course of the epidemic, but only to use a limited number of daily tests to maximal advantage towards minimising the number of infected individuals attending a local essential facility. This hinges principally upon the realisation that after a negative test, the probability of being infected for a particular individual, $P(n)$ in Fig 1, raises gradually from zero, and only after a certain period of time, τ_I, overtakes that same probability for the average global population. Thus, a limited number of practicable tests per day are best used by removing from the testing sample individuals having been already tested within the preceding τ_I day period. To use a physical analogy, standard epidemiological models, e.g. of the S-I-R type, aim at tracking the evolution of the infection rate (which becomes analogous to the temperature of a physical system) over an entire population, with derived proposed interventions aimed at reducing the overall spread and duration of the epidemic, see for example [17] for a treatment of an optimal sampling and testing strategy for the general population. Hence, such models can be compared to working within the microcanonical ensamble, where the evolution of the temperature of the system can be modeled. The approach we have presented is analogous to working within the canonical ensamble, where the temperature is fixed through contact to an external thermal bath which is assumed as a restriction on the smaller system whose dynamics are to be explored, under the restriction of a temperature (in our analogy the infection rate) which is determined and imposed by external agencies.

The probabilistic model presented and the numerical simulations shown support the validity and generality of the approach developed, whenever the assumptions upon which the model was constructed are valid. Whenever reality might deviate substantially from such

assumptions, the results presented would be invalidated. Chiefly, we have assumed a constant and well known duration of the infection for all individuals, $\tau_I = 15$ days. The probabilistic approach is robust towards a distribution of infection periods present in the population, provided this distribution is not skewed, and the sampling exclusion period for individuals is adjusted to the mean τ_I present in the overall population, which must be known. The monotonously decreasing slope of $P(n)$ in Fig 1 results in the broad minimums in IPD seen in Figs 2 and 3, where the IPD numbers drop rapidly at first with increasing n_{ex} values, and then converge to the optimal values with little further change. This point is important as it implies that taking n_{ex} values smaller than the mean of τ_I will result in much larger excess IPD values than taking n_{ex} thresholds larger than mean τ_I.

In the absence of PRC tests one might have to rely upon less secure antigen testing, where a certain false negative rate might be expected. This same situation will arise if (when?) mutations reduce the reliability of the original PCR tests, developed with the original viral variant in mind. The adjustment to the approach presented in this case is fairly straight forward, as the objective is to identify pre-symptomatic and asymptomatic individuals. If a 50% false negative rate is present in the testing procedure available, twice as many test are required to identify the sought after individuals, i.e. n_{ex} should be reduced by a factor of 2. In general $n_{ex} = \tau_I/(1 - F_N)$, where F_N is the false negative rate of the available test. Any false positive rate present in the testing used will clearly reduce the number of assisting personnel, but will not modify the optimal strategy presented.

Asides form the points already mentioned, a further caveat lies in the assumption of the small target population being a fair sample of the overall one, where critical parameters such as I_f and τ_I are generally measured. It is clear that the exact optimal testing strategy will depend upon the many details of exactly how the grounding assumptions of the model fail in a real situation, in ways which lie beyond the scope of this first presentation of the problem.

Conclusions

We here develop an optimal testing strategy designed for minimising the number of infected asymptomatic persons present during a pandemic in an essential workspace setting, under the constraint of a fixed number of tests per day. We stress the potential of the method in that overall IPD values can drop by a factor of 2 or more, compared to a simple fully random strategy at 10% daily test fractions, simply by choosing carefully the daily testing sample, at fixed daily test numbers.

The infected population substitution timescale, which corresponds to the infected period for the simple case where all infected individuals remain infected/infectious for exactly the same constant period of time, τ_I, is identified as the critical parameter of such an optimisation. This raises two important points i) The optimal use of a finite number of high-quality tests is to perform a random sampling of the present population from which members having been tested during the preceding τ_I days are excluded, and which is augmented if needed, by including those members not having been tested during the preceding $\tau_I - 1$ days, and so on, until the fixed number of tests per day can be randomly chosen, i.e. until the testing sample is larger than the number of daily tests. ii) It is crucial to have a more accurate and population weighted estimate of τ_I, which surely is only a universal constant to a very first approximation.

Supporting information

S1 File.
(C)

S2 File.
(PDF)

Acknowledgments

The authors acknowledge the constructive criticism of three anonymous referees as important towards reaching a more complete, clear and balanced version of our manuscript.

Author Contributions

Conceptualization: X. Hernandez, S. Valentinotti.

Formal analysis: X. Hernandez.

Investigation: X. Hernandez.

Methodology: X. Hernandez.

Software: X. Hernandez.

Writing – original draft: X. Hernandez, S. Valentinotti.

Writing – review & editing: X. Hernandez, S. Valentinotti.

References

1. Petersen I., Andrews P. Three Quarters of People with SARS-CoV-2 Infection are Asymptomatic: Analysis of English Household Survey Data. Clin. Epidemiol. 2020; 12:1039–1043. https://doi.org/10.2147/CLEP.S276825 PMID: 33116898
2. Buitrago-Garcia D, Egli-Gany D, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLOS MEDICINE. 2020; 100346. https://doi.org/10.1371/journal.pmed.1003346 PMID: 32960881
3. European Centre for Disease Prevention and Control. Transmission of Covid-19. https://www.ecdc.europa.eu/en/covid-19/latest-evidence/transmission. Last updated Sep. 7, 2020. Accessed December 22, 2020.
4. European Centre for Disease Prevention and Control. COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK. Stockholm: ECDC; 2020.
5. Weinberg C. Editorial: Making the Best Use of Test Kits for COVID-19. American Journal of Epidemiology 2020; 189 (5) 363–364. https://doi.org/10.1093/aje/kwaa080 PMID: 32378722
6. Attipoe-Dorcoo S, Delgado R, Gupta A, Bennet J, Oriol NE, Jain SH. Mobile health clinic model in the COVID-19 pandemic: lessons learned and opportunities for policy changes and innovation. International Journal for Equity in Health. 2020; 19(1): 1–5. https://doi.org/10.1186/s12939-020-01175-7 PMID: 32429920
7. Jefatura de Gobierno GACETA OFICIAL DE LA CIUDAD DE MEXICO. 2022; Vigesima Primera Epoca, No. 498 Bis
8. Grassly N, Pons-Salort M, et al. Comparison of molecular testing strategies for COVID-19 control: a mathematical modelling study. Lancet Infect Dis 2020; 20: 1381–1389. https://doi.org/10.1016/S1473-3099(20)30630-7 PMID: 32822577
9. Saidani M., Kim H., & Kim J. Designing optimal COVID-19 testing stations locally: A discrete event simulation model applied on a university campus. PloS one. 2021; 16(6), e0253869 https://doi.org/10.1371/journal.pone.0253869 PMID: 34185796
10. Fiore VG, DeFelice N, Glicksberg BS, Perl O, Shuster A, Kulkarni K, et al. Containment of COVID-19: Simulating the impact of different policies and testing capacities for contact tracing, testing, and isolation. PLoS ONE. 2021; 16(3): e0247614. https://doi.org/10.1371/journal.pone.0247614 PMID: 33788852
11. Lyng GD, Sheils NE, Kennedy CJ, Griffin DO, Berke EM. Identifying optimal COVID-19 testing strategies for schools and businesses: Balancing testing frequency, individual test technology, and cost. PLoS ONE. 2021; 16(3): e0248783. https://doi.org/10.1371/journal.pone.0248783 PMID: 33764982

12. Lamé G, Simmons RK. From behavioural simulation to computer models: how simulation can be used to improve healthcare management and policy. BMJ Simulation and Technology Enhanced Learning. 2020; 6(2): 95–102. https://doi.org/10.1136/bmjstel-2018-000377

13. Currie CS, Fowler JW, Kotiadis K, Monks T, Onggo BS, Robertson DA, et al. How simulation modelling can help reduce the impact of COVID-19. Journal of Simulation. 2020;1–15. https://doi.org/10.1080/17477778.2020.1751570

14. Johns Hopkins university & Medicine, Coronavirus Resource Center. Mortality Analyses. http://coronavirus.jhu.edu/data/mortality Last updated Tuesday December 22, 2020. Accessed December 22, 2020.

15. Adams K. Expandimg Covid-19 Testing: Mathematial Guidelines for the Optimal Sample Pool Size Given Positive Test Rate. medRxiv 2020.05.21.20108522.

16. Bogere N, Bongomin F, Katende A, Ssebambulidde K, Ssengooba W, Ssenfuka H, et al. Performance and cost-effectiveness of a pooled testing strategy for SARS-CoV-2 using real-time polymerase chain reaction in Uganda. Int J Infect Dis. 2021 Oct 28; 113:355–358. https://doi.org/10.1016/j.ijid.2021.10.038 PMID: 34757007

17. Deckert A, Anders S, de Allegri M, Nguyen HT, Souares A, McMahon S, et al. Effectiveness and cost-effectiveness of four different strategies for SARS-CoV-2 surveillance in the general population (CoV-Surv Study): a structured summary of a study protocol for a cluster-randomised, two-factorial controlled trial. Trials. 2021 Jan 8; 22(1):39. https://doi.org/10.1186/s13063-020-04982-z PMID: 33419461