# "No test is better than a bad test": Impact of diagnostic uncertainty in mass testing on the spread of COVID-19

Nicholas Gray[1*◕], Dominic Calleja[1◕], Alexander Wimbush[1◕], Enrique Miralles-Dolz[1], Ander Gray[1], Marco De-Angelis[1], Elfriede Derrer-Merk[1], Bright Uchenna Oparaji[1], Vladimir Stepanov[1], Louis Clearkin[2], Scott Ferson[1]

**1** Institute for Risk and Uncertainty, University of Liverpool, United Kingdom
**2** Wirral & Liverpool University Teaching Hospitals, United Kingdom

◕These authors contributed equally to this work.
* nickgray@liverpool.ac.uk, covid19@riskinstitute.uk

## Abstract

Testing is viewed as a critical aspect of any strategy to tackle the COVID-19 pandemic. Much of the dialogue around testing has concentrated on how countries can scale up capacity, but the uncertainty in testing has not received nearly as much attention beyond asking if a test is accurate enough to be used. Even for highly accurate tests, false positives and false negatives will accumulate as mass testing strategies are employed under pressure, and these misdiagnoses could have major implications on the ability of governments to suppress the virus. The present analysis uses a modified SIR model to understand the implication and magnitude of misdiagnosis in the context of ending lock-down measures. The results indicate that increased testing capacity alone will not provide a solution to lock-down measures in the UK. The progression of the epidemic and peak infections is shown to depend heavily on test characteristics, test targeting, and prevalence of the infection. Antibody based immunity passports are rejected as a solution to ending lockdown, as they can put the population at risk if poorly targeted. Similarly, mass screening for active viral infection may only be beneficial if it can be sufficiently well targeted, otherwise reliance on this approach for protection of the population can again put them at risk. A well targeted active viral test combined with a slow release rate is a viable strategy for continuous suppression of the virus.

## Introduction

The UK government's COVID-19 epidemic management strategy has been influenced by epidemiological modelling conducted by a number of research groups [1, 2]. The analysis of the relative impact of different mitigation and suppression strategies has influenced the current approach. The "only viable strategy at the current time" is to suppress the epidemic with all available measures, including school closures and social distancing of the entire population [3]. These analyses have highlighted from the beginning that the eventual relaxation of lock-down measures would be problematic [3, 4]. Without a considered cessation of the suppression strategies the risk of a second wave becomes signficant, possibly of greater magnitude than the first as the SARS-CoV-2 virus is now endemic in the population [5, 6].

Although much attention has been focused on the number of tests being conducted [7, 8], not enough attention has been given to the issues of imperfect testing.

Whilst poorly performing tests have not been prominent in public discourse, evidence suggests they are epidemiologically significant. The failure to detect the virus in infected patients can be a significant problem in high-throughput settings operating under severe pressure [9]. Evidence suggest that this is indeed the case [10–13].

Everyone seems to agree that testing will be a pillar of whatever approach is employed to relax the current social distancing measures [14]. The public are rapidly becoming aware of the difference between the 'have you got it?' tests for detecting active cases, and the 'have you had it?' tests for the presence of antibodies, which imply some immunity to COVID-19. What may be less obvious is that these different tests need to maximise different test characteristics.

To be useful in ending the current social distancing measures, active viral tests need to maximise the sensitivity. How good the test is at telling you that you have the disease. High sensitivity reduces the chance of missing people who have the virus who may go on to infect others. There is an additional risk that an infected person who has been incorrectly told they do not have the disease, when in fact they do, may behave in a more reckless manner than if their disease status were uncertain.

The second testing approach, seeking to detect the presence of antibodies to identify those who have had the disease would be used in a different strategy. This strategy would involve detecting those who have successfully overcome the virus, and are likely to have some level of immunity (or at least reduced susceptibility to more serious illness if they are infected again), so are relatively safe to relax their personal social distancing measures. This strategy would require a high test specificity, aiming to minimise how often the test tells someone they have had the disease when they haven't [15]. A false positive tells people they have immunity when they don't, which is even worse than when people are uncertain about their viral history.

## Evidence testing is flawed

The successes of South Korea, Singapore, Taiwan and Hong Kong in limiting the impact of the SARS-CoV-2 virus has been attributed to their ability to deploy widespread testing, with digital surveillance, and impose targeted quarantines in some cases [9]. This testing has predominantly been based on the use of reverse transcription polymerase chain reaction (RT-PCR) testing. During the 2009 H1N1 pandemic the rapid development of high sensitivity PCR assay were employed early with some success in that global pandemic [21]. These tests, when well targeted, clearly provide a useful tool for managing and tracking pandemics.

These tests form the basis of much of the research into the incidence, dynamics and comorbidities of SARS-CoV-2, but few, if any, of these studies give consideration to the impact of false test results [22–26]. Increasing reliance on lower-sensitivity tests to address capacity concerns is likely to make available data on confirmed cases more difficult to accurately utilise [21]. It may be the case that false test results contribute to some of the counter-intuitive disease dynamics observed [27].

There is evidence that both active infection [28–32] and antibody [33–35] tests lack perfect sensitivity and specificity even in best-case scenarios. Alternative screening methods such as chest x-rays may be found to have high sensitivity based on biased data [36] or may simply perform poorly even compared to imperfect RT-PCR tests [31]. The Foundation for Innovative New Diagnostics (FIND) conducted an independent evaluation of five RT-PCR tests which scored highly out of 17 candidate tests on criteria such as regulatory status and availability [37]. Even ideal laboratory conditions can produce a specificity which could be as low as 90%, and the practical specificity is likely to be lower still.

## Introduction to test statistics: What makes a 'good' test?

In order to answer this question there are a number of important statistics:

- **Sensitivity** $\sigma$ - Out of those who actually have the disease, that fraction that received a positive test result.
- **Specificity** $\tau$ - Out of these who did not have the disease, the fraction that received a negative test result.

The statistics that characterise the performance of the test are computed from a confusion matrix (Table 1). We test $n_{infected}$ people who have COVID-19, and $n_{healthy}$ people who do not have COVID-19. In the first group, $a$ people correctly test positive and $c$ falsely test negative. Among healthy people, $b$ will falsely test positive, and $d$ will correctly test negative.

| | Infected | Not Infected | Total |
|---|---|---|---|
| Tested Positive | $a$ | $b$ | $a+b$ |
| Tested Negative | $c$ | $d$ | $c+d$ |
| Total | $a+c=n_{infected}$ | $b+d=n_{healthy}$ | $N$ |

**Table 1.** Confusion matrix

From this confusion matrix the sensitivity is given by (1) and the specificity by (2).

$$\sigma = \frac{a}{n_{infected}} \tag{1}$$

$$\tau = \frac{d}{n_{healthy}}. \tag{2}$$

Sensitivity is the ratio of correct positive tests to the total number of infected people involved in the study characterising the test. The specificity is the ratio of the correct negative tests to the total number of healthy people. Importantly, these statistics depend only on the test itself and do not depend on the population the test is intended to be used upon.

Computing the statistics require a definitive way to determine the true viral status of a patient; a so-called *gold standard*. If there is doubt about in which column a patient falls, the confusion matrix cannot be constructed. When novel tests are employed the confusion matrix can be very challenging to rigorously assess in the midst of a fast moving epidemic [16–18].

When the test is used for diagnostic purposes, the characteristics of the population being tested become important for interpreting the test results. To interpret the diagnostic value of a positive or negative test result the following statistics must be used:

- **Prevalence** $p$ - The proportion of people in the target population that have the disease tested for.
- **Positive Predictive Value** $PPV$ - How likely one is to have the disease given a positive test result.
- **Negative Predictive Value** $NPV$ - How likely one is to *not* have the disease, given a negative test result.

The $PPV$ and $NPV$ depend on the prevalence, and hence depend on the population you are focused on. This may be the UK population, a sub population with COVID-19 compatible symptoms, or any other population you may wish to target. The $PPV$ and $NPV$ can then be calculated using Bayes' rule:

$$PPV = \frac{p\sigma}{p\sigma + (1-p)(1-\tau)}, \tag{3}$$

$$NPV = \frac{\tau(1-p)}{\tau(1-p) + (1-\sigma)p}. \tag{4}$$

To improve the diagnostic performance of tests they are often repeated to increase the aggregate $PPV$ or $NPV$. To do this an assumption of independence between the two tests needs to be made. This assumption could be questionable in some circumstances. For instance, it would be questionable if samples are analysed at the same time in the same lab by the same technician, or if the same method for extracting the sample from the patient is employed, it may be unsuccessful at detecting virus for the same reason. A plethora of other possible errors are imaginable. Many of these errors may be truly random, and independent, but many may not be so the independence assumption may be weakly justified.

The rapid development and scaling of new diagnostic systems invites error, particularly as labs are converted from other purposes and technicians are placed under pressure, and variation in test collection quality, reagent quality, sample preservation and storage, and sample registration and provenance. Assessing the magnitude of these errors on the performance of tests is challenging in real time. Point-of-care tests are not immune to these errors and are often seen as less accurate than laboratory based tests [19, 20].

### Why mass testing may be ineffectual

The prevalence of the disease matters. The $PPV$ can vary drastically for different populations with different prevalence. The idea that prevalence depends on the population may seem counterintuitive to some audiences. For example, if we were to select 100 people from a respiratory ward this week from any hospital in the UK, and 100 people from a street outside the building, what proportion of each population have COVID-19? If one tests both populations with the same test and found positives in each population, which would have the higher $PPV$?

To illustrate the impact of prevalence on $PPV$, for a test with $\sigma = \tau = 0.95$ if prevalence $p = 0.05$, then the $PPV \approx 0.48$. Figure 1 shows why, for 1000 test subjects there will be similar numbers of true and false positives even with high sensitivity and specificity of 95%. In contrast, using the same tests on a sample with a higher prevalence $p = 0.5$ we find the $PPV = 0.8$, see Figure 2. Similarly, the $NPV$ is lower when the prevalence is higher.

The number of active cases exceeding the test capacity may not be the only discrepancy between the true cases and reported cases. The impact of uncertainty in testing may also be contributing to the discrepancy, even in the tested population. More testing will not reduce this uncertainty. The director of the WHO suggests that testing is a crucial part of any strategy [38], but even testing the entire country every day would not give an accurate tally of infections.

## SIR model with testing

To explore the effect of imperfect testing on the disease dynamics when strategies are employed to relax the current social distancing measures the SIR model described in the

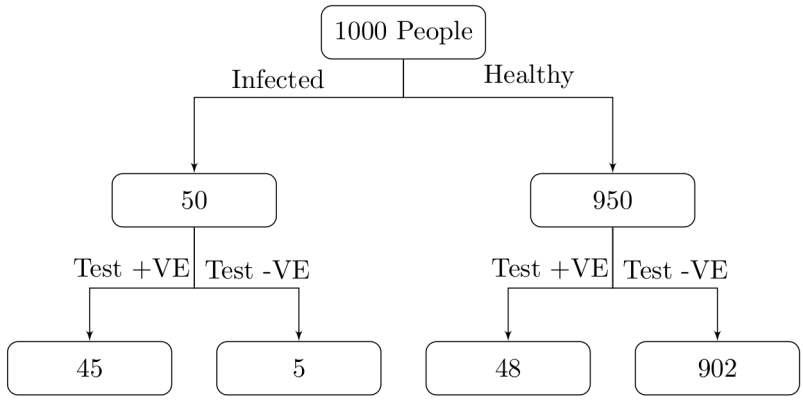


**Fig 1.** If the prevalence of a disease amongst those being tested is 0.05 then with $\sigma = \tau = 0.95$ the number of false positives will outnumber the true positives, resulting in a low $PPV$ of around 0.48.

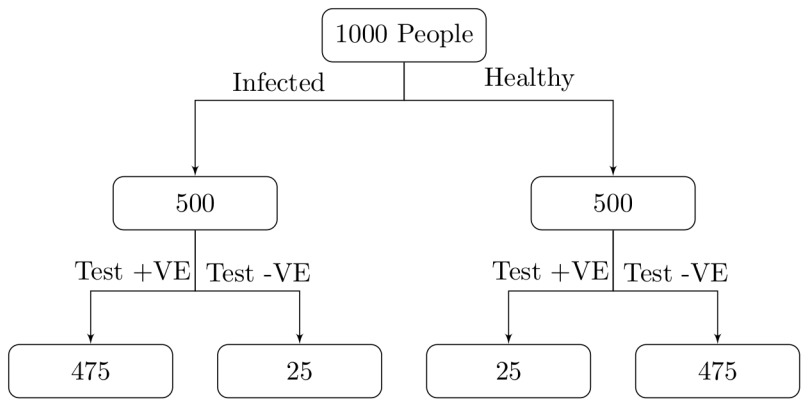


**Fig 2.** If the prevalence of a disease amongst those being tested is 0.50 then with $\sigma = \tau = 0.95$ the number of true positives will outnumber the number of false positives, resulting in a high $PPV$ of 0.95.

supplementary material was modified. Three new classes were added to the model, the first is a quarantined susceptible state, $Q_S$, the second is a quarantined infected state, $Q_I$, and the third is people who have recovered but are in quarantine, $Q_R$.

To model the current lock-down, the model evaluations begin with a majority of the population in the $Q_S$ (quarantined but susceptible) state. Whilst in this state the transmission rate of the disease is totally suppressed. The model evaluates each day's average population-level state transitions. There are two possible tests that can be performed:

- An active virus infection test that is able to determine whether or not someone is currently infectious. This test is performed on some proportion of the un-quarantined population ($S + I + R$). It has a sensitivity of $\sigma_A$ and a specificity of $\tau_A$.
- An antibody test that determines whether or not someone has had the infection in the past. This is used on the fraction of the population that is currently in quarantine but not infected ($Q_S + Q_R$) to test whether they have had the disease or not. This test has a sensitivity of $\sigma_B$ and a specificity of $\tau_B$.

These two tests are used on some of those eligible for testing each day limited by the

test capacity, $\rho$ and $\phi$ respectively. A person (in any category) who tests positive in an active virus test transitions into the (corresponding) quarantine state, where they are unable to infect anyone else. A person, in $Q_S$ or $Q_R$, who tests positive in an antibody test transitions to $S$ and $R$ respectively.

For this parameterisation the impact of being in the susceptible quarantined state, $Q_S$, makes an individual insusceptible to being infected. Similarly, being in the infected quarantined state, $Q_I$, individuals are unable to infect anyone else. In practicality there is always leaking, no quarantine is entirely effective, but for the sake of exploring the impact of testing uncertainty these effects are neglected from the model. The participation in infection propagation of individuals in either quarantine state are idiosyncratic, and on average are assumed to be negligibly small for the sake of this analysis.

If the tests were almost perfect, then we can imagine how the epidemic would die out very quickly by either widespread infection or antibody testing with a coherent management strategy. A positive test on the former and the person is removed from the population, and positive test on the latter and the person, unlikely to contract the disease again, can join the population .

More interesting are the effects of incorrect test results on the disease dynamics. If someone falsely tests positive in the antibody test, they enter the susceptible state. Similarly, if an infected person receives a false negative for the disease they remain active in the infected state and hence can continue the disease propagation and infect further people.

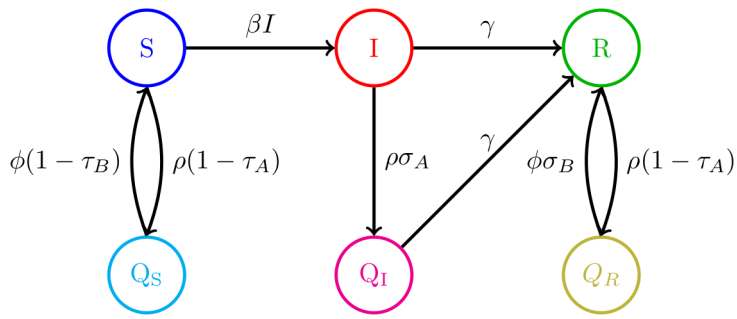


**Fig 3.** SIR Model used to simulated the effect of mass testing to leave quarantine.

## What part will testing play in relaxing current social distancing measures?

In order to explore the possible impact of testing strategies on the relaxation of current social distancing measures several scenarios have been analysed. These scenarios are illustrative of the type of impact, and the likely efficacy of a range of different testing configurations.

- **Immediate end to social distancing scenario**: This baseline scenario is characterised by a sudden relaxation of the current social distancing measures.
- **Immunity passports scenario**: A policy that has been discussed in the media [18, 39, 40]. Analogous to the International Certificate of Vaccination and Prophylaxis, antibody based testing would be used to identify those who have some level of natural immunity.

- **Incremental relaxation scenario**: A phased relaxation of the government's social distancing advice is the most likely policy that will be employed. To understand the implications of such an approach this scenario has explored the effect of testing capacity and test performance on the possible disease dynamics under this type of policy. Under the model parameterisation this analysis has applied an incremental transition rate from the $Q_S$ state to the $S$ state, and $Q_R$ to $R$.

Whilst the authors are sensitive to the sociological and ethical concerns of any of these approaches [41, 42], the analysis presented is purely on the question of efficacy.

## Immediate end to lock-down scenario

Under the baseline scenario, characterised by the sudden and complete cessation of the current social distancing measures, we explored the impact of infection testing. Under this formulation the initial conditions of the model in this scenario is that the all of the population in $Q_S$ transition to $S$ in the first iteration.

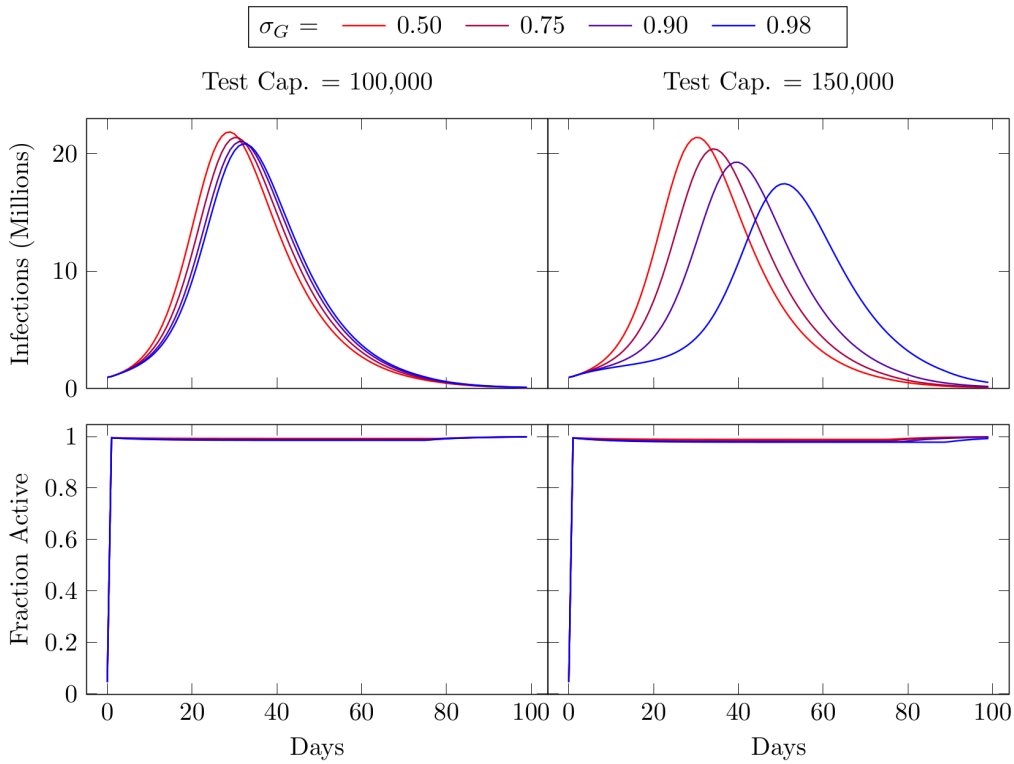


**Fig 4.** A comparison of different infection test sensitivities $\sigma_G$ shown from red to blue. Two different infection test capacities are considered. Left: test capacity $= 1 \times 10^5$. Right: test capacity $= 1.5 \times 10^5$. Top: The number of infected individuals ($I + Q_I$ population) over 100 days. Bottom: The proportion of the population that has been released from quarantine ($S + I + R$ population) over 100 days.

As would be expected the model indicates the second wave is an inevitability and as many as 20 million people could become infected within 30 days, figure 4. To illustrate the sensitivity of the model to testing scenarios an evaluation was conducted with a range of infection test sensitivities, from 50% (i.e of no diagnostic value) to 98%. The specificity of these tests has a negligible impact on the disease dynamics. A false

positive test result would mean people are unnecessarily removed from the susceptible population, but the benefit of a reduction in susceptible population is negligibly small. It's also very likely the infection testing would be heavily biased toward symptomatic carriers, where the prevalence of the disease is high so fewer false positives would be expected.

Two evaluations have been conducted. The first using the stated government goal of 100,000 tests per day (left graphs in figure 4). It remains unclear whether this aim is feasible, or if this testing capacity would include both forms of tests (antibody and active virus). The second evaluation looks at a very optimistic case where we could conduct as many as 150,000 tests per day (right graphs in figure 4). The authors draw no conclusions about the feasibility of achieving these levels. However the authors do wish to encourage caution that with a capacity for testing of the order targeted by the UK government, testing in isolation is not sufficient to allow any rapid cessation of the current social distancing measures without a resurgence of the virus. This caution is irrespective of test performance, even very good tests with a sensitivity of 98%, and effective isolation of cases that have tested positive, the outcome is broadly invariant.

## Immunity passports scenario

The immunity passport is an idiom describing an approach to the relaxation of the current social distancing measures that focuses heavily on antibody testing. Wide-scale screening for antibodies in the general population promises significant scientific value, and targeted antibody testing is likely to have value for reducing risks to NHS and care-sector staff, and other key workers who will need to have close contact with COVID-19 sufferers. The authors appreciate these other motivations for the development and roll-out of accurate antibody tests. This analysis however focuses on the appropriateness of this approach to relaxing current social distancing measures by mass testing the general population. Antibody testing has been described as a 'game-changer' [43]. Some commentators believe this could have a significant impact on the relaxation of social distancing measures [40], but others note that there are severe ethical, logistical and medical concerns which need to be resolved before antibody testing could support a strategy such as this [44].

Much of the discussion around antibody testing in the media has focused on the performance and number of these tests. The efficacy of this strategy however is far more dependent on the prevalence of antibodies in the general population. Without wide-scale antibody screening it's impossible to know the prevalence of antibodies in the general population, so there is scientific value in such an endeavour. However, the prevalence is the dominant factor to determine how efficacious antibody screening would be for relaxing social distancing measures.

Presumably, only people who test positive for antibodies would be allowed to leave quarantine. The more people in the population with antibodies, the more people will get a true positive, so more people would be correctly allowed to leave quarantine (under the paradigms of an immunity passport).

The danger of such an approach is the false positives. We demonstrate the impact of people reentering the susceptible population who have no immunity. We assume their propensity to contract the infection is the same as those without this the false sense of security a positive test may engender. On an individual basis, and even at the population level, behavioural differences between those with false security from a positive antibody test, versus those who are uncertain about their viral history could be significant. The model parametrisation here does not include this additional confounding effect.

To simulate the prevalence of antibodies in the general population the model is preconditioned with different proportions of the population in the $Q_S$ and $Q_R$ states.

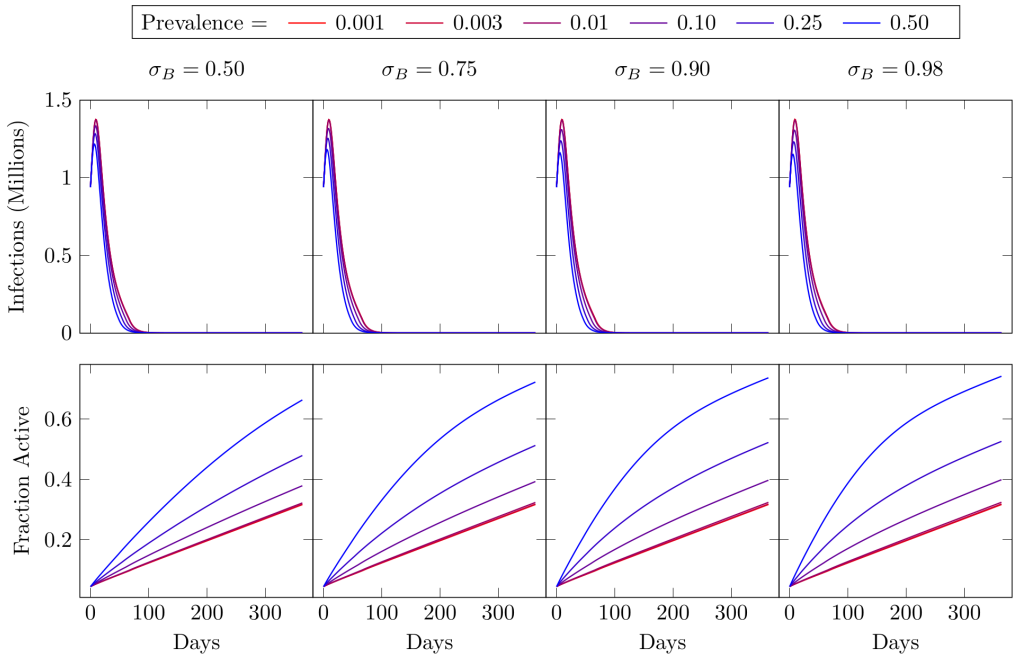


**Fig 5.** A comparison of different antibody test sensitivities $\sigma_B$, with varying levels of prevalence. Top: The number of infected individuals ($I + Q_I$ population) over one year. Bottom: The proportion of the population that has been released from quarantine ($S + I + R$ population) over one year.

This is analogous to the proportion of people that are currently in quarantine who have either had the virus and developed some immunity, and the proportion of the population who have not contracted the virus and have no immunity. Of course the individuals in these groups do not really know their viral history, and hence would not know which state they begin in. The model evaluations explore a range of sensitivity and specificities for the antibody testing. These sensitivity and specificities, along with the capacity for testing, govern the transition of individuals from $Q_R$ to $R$ (true positive tests), and from $Q_S$ to $S$ (false positive tests).

Figures 5 and 6 show the results of the model evaluations. The top row of each figure corresponds to the number of infections in time, the bottom row of each figure is the proportion of the population that are released from quarantine and hence are now in the working population. Maximising this rate of reentry into the population is of course desirable, and it is widely appreciated that some increase in the numbers of infections is unavoidable. The desirable threshold in the trade-off between societal activity and number of infections is open to debate.

Each column corresponds to a different antibody test sensitivity in figure 5 as titled. The specificity for each test in these evaluations was fixed to 90%. In figure 6 each column corresponds to a different antibody test specificity as titled. The sensitivity for each test in these evaluation was fixed to 95%. For all model evaluations in figures 5 and 6 we modelled a continuing and constant ability to conduct targeted active virus testing which continued to remove individuals from the infected population. The infection testing continued throughout each model run with a fixed capacity of 10,000 tests per day, similar to the number of unique individuals that are currently being tested.

Each of the graphs in the two figures shows the effect of different prevalences of antibodies in the population. To be clear, this is the proportion of the population that has contracted the virus and recovered but are in quarantine. Sir Patrick Vallance, the

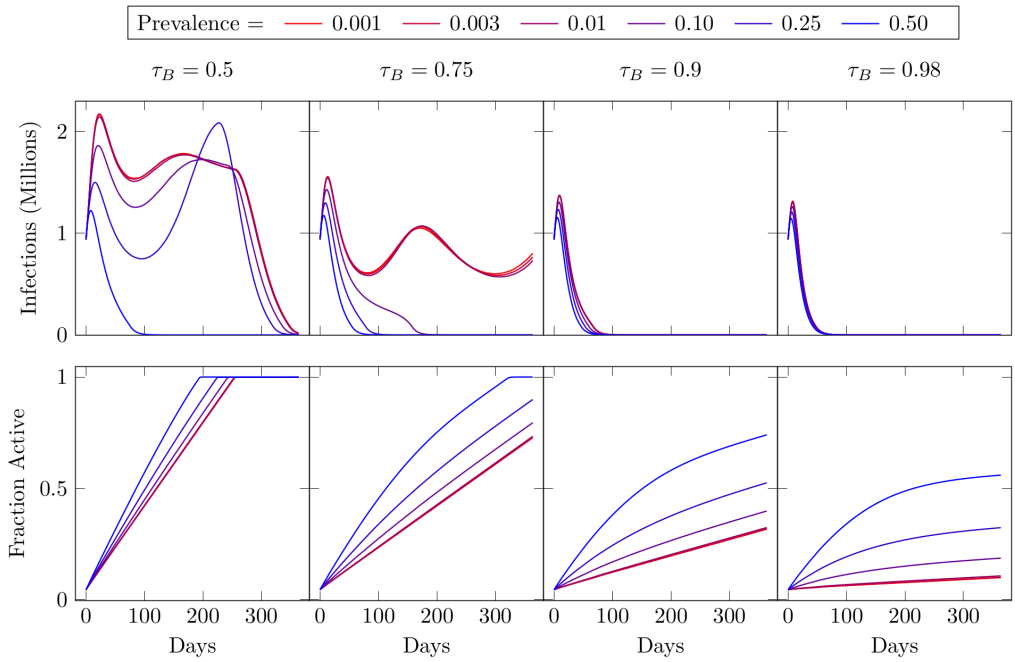


**Fig 6.** A comparison of different antibody test specificities $\tau_B$ shown from left to right, with varying levels of prevalence shown from red to blue. Top: The number of infected individuals ($I + Q_I$ population) over one year. Bottom: The proportion of the population that has been released from quarantine ($S + I + R$ population) over one year.

UK Government Chief Scientific Advisor, in the daily press briefing on the 9 April 2020 stated his belief that this prevalence is likely to be less than 10%, possibly much less. The analysis has explored a range for prevalence from 0.1% to 50%. Figure 5 explores the impact of a variation in sensitivity, from a test with 50% sensitivity (i.e no diagnostic value) to tests with a high sensitivity of 98%. It can be seen, considering the top half of the graphs, that the sensitivity of the test has no discernible impact on the number of infections. The prevalence entirely dominates. This is possibly counter intuitive, but as was discussed above, even a highly accurate test produces a very large number of false positives when prevalence is low. In this case that would mean a large number of people are allowed to re-enter the population, placing them at risk, with a false sense of security that they have immunity.

The bottom row of figure 5 shows the proportion of the entire population leaving quarantine over a year of employing this policy. At low prevalence there is no benefit to better performing tests. This again may seem obscure to many readers. If you consider the highest prevalence simulation, where 50% of the population have immunity, higher sensitivity tests are of course effective at identifying those who are immune, and gets them back into the community much faster. However this is not the case currently in the UK because, as Sir Patrick stated, the prevalence of antibodies is likely to be very low at least during the lock-down.

A more concerning story can be seen when considering the graphs in figure 6. Now we consider a range of antibody test specificities. Going from 50% (no value in ruling people out) to 98%. When the prevalence is low, a lower specificity of 75% not only leads to an initial large increase in the number of infections, but also, if employed throughout the year would lead to repeated peaks. This is because the active virus testing would still be employed along side the antibody testing. Falsely-diagnosed susceptible people leaving quarantine leads to a sharp rise in the number of infections.

As the prevalence of virus in the non-quarantined population increases the active virus testing becomes more effective and subdues the rise in infections because the testing is more targeted on active virus cases. This would be followed by additional waves as further false positive tests for antibodies are observed. The number of people in quarantine with antibodies declines over the length of the simulation, so naturally the prevalence of immunity in the quarantined population declines. As the prevalence declines the $NPV$ of the test declines.

When we consider the bottom half of figure 6 and look at the impact on the proportion of the population able to leave quarantine, unlike previously, the number of false positives dominates when there is a lower specificity. So there are many more people leaving the quarantine, even when the prevalence is very low (0.1%). This may be desirable to some who favour increasing economic and social activity, but it is of course at the cost of further infections. Decision makers and the public need to be aware of the trade-off being made. The dangers of neglecting uncertainties in medical diagnostic testing are pertinent to this decision [45], particularly if immunity passports become prominent in the strategy to end the current social distancing measures.

## Incremental relaxation scenario

At this point, some form of incremental relaxation of the current government social distancing advice seems highly likely. This could take many forms, it could be an incremental restoration of certain activities such as school openings, permission for the reopening of some businesses, the relaxation of the stay-at-home messaging, etc. Under the parameterisation chosen for this analysis the model is not sensitive to any particular policy change. We consider a variety of rates of phased relaxations to the current quarantine. To model these rates we consider a weekly incremental transition rate from $Q_S$ to $S$, and $Q_R$ to $R$. In figure 7, three weekly transition rates have been applied: 1%, 5% and 10% of the quarantined population. Whilst in practice the rate is unlikely to be uniform as decision makers would have the ability to update their timetable as the impact of relaxations becomes apparent, it is useful to illustrate the interaction of testing capacity and release rate.

The model simulates these rates of transition for a year, with a sensitivity and specificity of 90% for active virus tests. The specifics of all the runs are detailed in table 2. Figure 7 shows five analyses, with increasing capacity for the active virus tests. In each, the 3 incremental transition rates are applied with a range of disease prevalences in the population being tested. The $PPV$, as discussed above, has a greater dependence on the prevalence (at lower values) in the tested population than it does on the sensitivity of the tests, the same is true of the specificity and the $NPV$.

It is important to notice that higher test capacities cause a higher peak of infections for the 10% quarantine release rate. This has a counterintuitive explanation. When there is the sharpest rise in the susceptible population (i.e., high rate of transition), the virus rapidly infects a large number of people. When these people recover after two weeks they become immune and thus cannot continue the spread of the virus. However, when the infection testing is conducted with a higher capacity up to 120,000 units per day, these tests transition some active viral carriers into quarantine, so the peak is slightly delayed providing more opportunity for those released from quarantine later to be infected, leading to higher peak infections. This continues until the model reaches effective herd immunity after which the number of infected in the population decays very quickly. Having higher testing capacities delays but actually worsens the peak number of infections.

At 10% release rate, up to a capacity of testing of 120,000 these outcomes are insensitive to the prevalence of the disease in the tested population . This analysis indicates that the relatively fast cessation of social distancing measures and stay-home

advice would lead to a large resurgence of the virus. Testing capacity of the magnitude stated as the goal of the UK government would not be sufficient to flatten the curve in this scenario.

At the rate of 5% of the population in lock-down released incrementally each week the infection peak is suppressed compared to the 10% rate. The number of infections would remain around this level for a significantly longer period of time, up to 6 months. There is negligible impact of testing below a capacity of 50,000 tests. However if the test capacity were 80,000 tests, at a quarantine release rate of 5% the duration of the elevated levels of infections would be reduced, reducing the length of necessary wide-scale social distancing. This effect is only observed with the more targeted tests, where a prevalence of the disease in the targeted population is over 30%. Any less well targeted testing and the testing would have a negligible impact compared to the untested scenario.

The 1% release rate scenario indicates that a slow release by itself is sufficient to lower peak infections, but extends the duration of elevated infections. The first graph of the top row in figure 7 shows that the slow release rate causes a plateau at a significantly lower number of infections compared to the other release rates. Poorly targeted tested at capacities less than 100,000 show similar consistent levels of infections. However, with a targeted test having a prevalence of 30% or more, the 1% release rate indicates that even with 50,000 tests per day continuous suppression of the infection may be possible.

| Model Parameters | | | | Initial Population split | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\sigma_A$ | $\tau_A$ | $\beta$ | $\gamma$ | $Q_S$ | $S$ | $Q_I$ | $Q_I$ | $Q_R$ | $R$ |
| 0.9 | 0.9 | 0.32 | 0.1 | 0.95 | 0.034 | 0.004 | 0.01 | 0.001 | 0.001 |

**Table 2.** Fixed parameters used for Figure 7 analysis.

# Conclusions

This analysis does support the assertion that a bad test is worse than no tests, but a good test is only effective in a carefully designed strategy. More is not necessarily better and over estimation of the test accuracy could be extremely detrimental.

This analysis is not a prediction; the numbers used in this analysis are estimates, and therefore, when such policies are devised and implemented this analysis would need to be repeated with more up-to-date numerical values. As such, the authors are not drawing firm conclusions about the absolute necessary capacity of tests. Nor do they wish to make specific statements about the necessary sensitivity or specificity of tests or the recommended rate of release from quarantine. The authors do, however, propose some conclusions that would broadly apply to the present situation, and therefore believe they should be considered by policy makers when designing strategies to tackle COVID-19.

- Diagnostic uncertainty can have a large effect on the epidemic dynamics of COVID-19 within the UK. And, sensitivity, specificity, and the capacity for testing alone are not sufficient to design effective testing procedures.

- Great caution should be exercised in the use of antibody testing. Under the assumption that the proportion of people in the UK who have had the virus is still low, it's unlikely antibody testing at any scale will significantly support the end of lock-down measures. And, the negative consequences of un-targeted antibody screening at the population level could cause more harm than good.

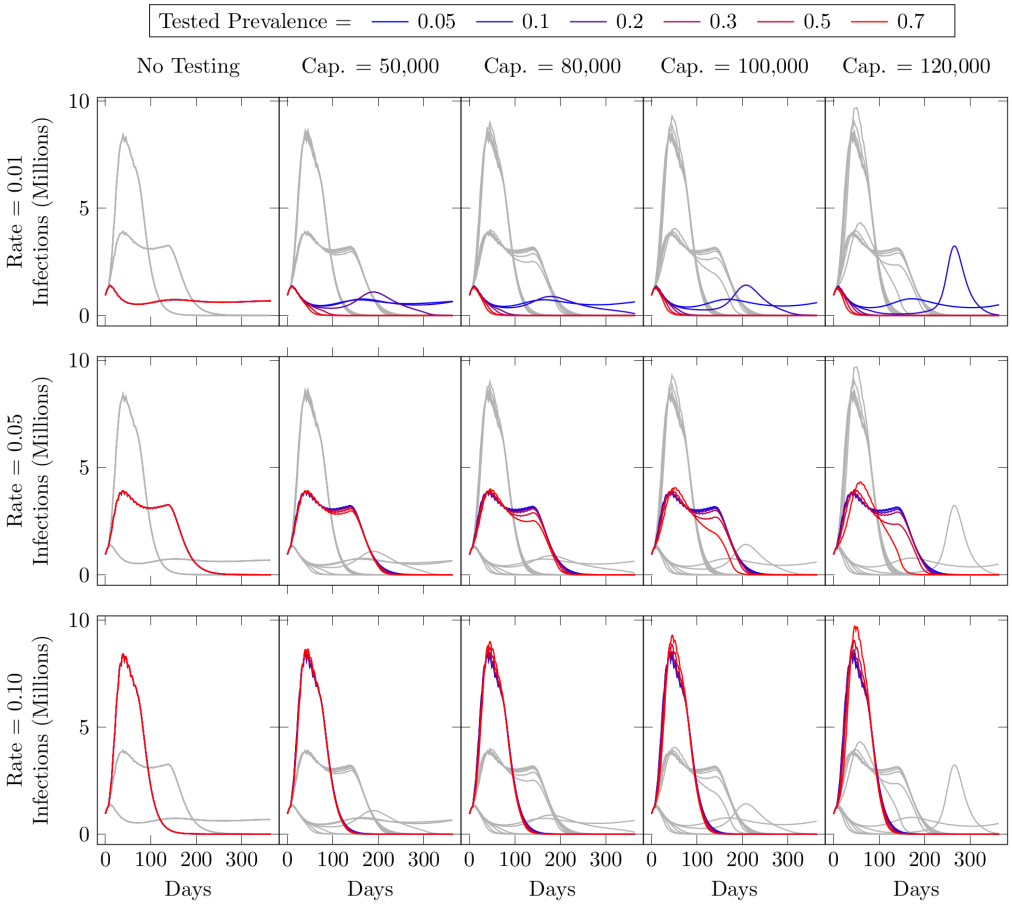


**Fig 7.** Total active infections each day over the year after relaxing lock-down, under different testing intensities (columns) and various epidemiologic conditions. The per-day testing capacity is varied across the five columns of graphs. Rate, the percentage of the initial quarantined population being released each week is varied among rows. The prevalence of infections in the tested population is varied among different colours. To facilitate comparison within each column of graphs, the gray curves show the results observed for other Rates and Prevalences with the same testing intensity.

- Antibody testing, with a high specificity may be very useful on an individual basis, it certainly has scientific value, and could reduce risk for key workers. But any belief that these tests would be useful to relax lock-down measures for the majority of the population is misguided. At best it is a distraction, at worst it could be dangerous.
- The incremental relaxation to lock-down measures, with all else equal, would significantly dampen the increase in peak infections, by 1 order of magnitude with a faster relaxation, and 2 orders of magnitude with a slower relaxation.
- The capacity for infection screening needs to be significantly increased if it is to be used to relax quarantine measures, but only if it is well targeted, for example through effective contact tracing. Untargeted mass screening would be ineffectual and may prolong the necessary implementation of lock-down measures.
- One interpretation of these results is that countries that had mass testing regimes early in the pandemic but had much lower case fatality rates may have been

reporting a large number of false positives.

- The results of this paper may explain what is being observed in nations such as Singapore as they continue to employ less-targeted mass testing and after a rapid cessation to their lock-down measures are now experiencing a second peak in infections [46].

## Acknowledgments

This work has been partially funded by the EPSRC IAA exploration award with grant number EP/R511729/1, EPSRC programme grant "Digital twins for improved dynamic design", EP/R006768/1, and the EPSRC and ESRC Centre for Doctoral Training in Quantification and Management of Risk and Uncertainty in Complex Systems and Environments, EP/L015927/1 .

## References

1. Department of Health and Social Care. COVID-19: government announces moving out of contain phase and
into delay phase; 2020. Available from: `https://www.gov.uk/government/news/COVID-19-government-announces-moving-out-of-contain-phase-and-into-delay`.

2. Department of Health and Social Care. Scaling up our testing programmes. Department of Health and Socal Care; 2020. April. Available from: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment{_}data/file/878121/coronavirus-COVID-19-testing-strategy.pdf`.

3. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team; 2020. doi:10.25561/77482.

4. Johnson B. PM address to the nation on coronavirus: 23 March 2020; 2020. Available from: `https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020`.

5. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health. 2020;2667(20):1–10. doi:10.1016/s2468-2667(20)30073-6.

6. Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. The Lancet. 2020;6736(20). doi:10.1016/S0140-6736(20)30746-7.

7. Horton R. Offline: COVID-19 and the NHS —"a national scandal". The Lancet. 2020;395(10229):1022. doi:10.1016/S0140-6736(20)30727-3.

8. Horton R. Offline: COVID-19—bewilderment and candour. The Lancet. 2020;395(10231):1178. doi:10.1016/S0140-6736(20)30850-3.

9. Sheridan C. Coronavirus and the race to distribute reliable diagnostics. Nature Biotechnology. 2020;38(April):379–391. doi:10.1038/d41587-020-00003-1.

10. Cummins E. Why the Coronavirus Test Gives So Many False Negatives; 2020. Available from: https://slate.com/technology/2020/04/coronavirus-testing-false-negatives.html.

11. Hao Q, Wu H, Wang Q. Difficulties in False Negative Diagnosis of Coronavirus Disease 2019: A Case Report. Infectious Diseases - Preprint. 2020; p. 1–12. doi:10.21203/RS.3.RS-17319/V1.

12. Krumholz HM. If You Have Coronavirus Symptoms, Assume You Have the Illness, Even if You Test Negative; 2020. Available from: https://www.nytimes.com/2020/04/01/well/live/coronavirus-symptoms-tests-false-negative.html.

13. Lichtenstein K. Are Coronavirus Tests Accurate?; 2020. Available from: https://www.medicinenet.com/script/main/art.asp?articlekey=228250.

14. Petherick A. Developing antibody tests for SARS-CoV-2. The Lancet. 2019;395(10230):1101–1102. doi:10.1016/S0140-6736(20)30788-1.

15. Grassly NC, Pons-salort M, Parker EPK, White PJ, Ainslie K, Baguelin M. Report 16 : Role of testing in COVID-19 control. 2020;(April):1–13.

16. Cohen J, Kupferschmidt K. Labs scramble to produce new coronavirus diagnostics. Science. 2020;367(6479):727–727. doi:10.1126/science.367.6479.727.

17. World Health Organisation. Medical Product Alert No3/2020: Falsified medical products, including in vitro diagnostics, that claim to prevent, detect, treat or cure COVID-19; 2020. Available from: https://www.who.int/news-room/detail/31-03-2020-medical-product-alert-n-3-2020.

18. Smyth C, Kennedy D, Kenber B. Britain has millions of coronavirus antibody tests, but they don't work; 2020. Available from: https://www.thetimes.co.uk/article/britain-has-millions-of-coronavirus-antibody-tests-but-they-don-t-work-j7

19. Green K, Graziadio S, Turner P, Fanshawe T, Allen J. Molecular and antibody point-of-care tests to support the screening , diagnosis and monitoring of COVID-19. Oxford COVID-19 Evidence Service; 2020. Available from: https://www.cebm.net/covid-19/molecular-and-antibody-point-of-care-tests-to-support-the-screening-diagn

20. David K. Point-of-Care Versus Lab-Based Testing: Striking a Balance; 2016. Available from: https://www.aacc.org/publications/cln/articles/2016/july/point-of-care-versus-lab-based-testing-striking-a-balance.

21. Jernigan DB, Lindstrom SL, Johnson JR, Miller JD, Hoelscher M, Humes R, et al. Detecting 2009 pandemic influenza a (H1N1) virus infection: Availability of diagnostic testing led to rapid pandemic response. Clinical Infectious Diseases. 2011;52(SUPPL. 1). doi:10.1093/cid/ciq020.

22. Qifang B, Yongsheng W, Shujiang M, Chenfei Y, Xuan Z, Zhen Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv Pre-Print. 2020;3099(20):1–9. doi:10.1016/S1473-3099(20)30287-5.

23. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine. 2020;382(13):1199–1207. doi:10.1056/NEJMoa2001316.

24. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. New England Journal of Medicine. 2020; p. 1–14. doi:10.1056/nejmoa2006100.

25. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy; 2020. Available from: `http://www.nature.com/articles/s41591-020-0883-7`.

26. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020 doi:10.1126/science.abb5793.

27. Korea Center for Disease Control and Prevention. Updates on COVID-19 in Republic of Korea; 2020. Available from: `https://www.cdc.go.kr/board/board.es?mid=a30402000000{&}bid=0030{&}act=view{&}list{_}no=366817{&}tag={&}nPage=1`.

28. Li Z, Yi Y, Luo X, Xiong N, Liu Y, Li S, et al. Development and Clinical Application of A Rapid IgM-IgG Combined Antibody Test for SARS-CoV-2 Infection Diagnosis. Journal of Medical Virology. 2020;(February). doi:10.1002/jmv.25727.

29. Yang Y, Yang M, Shen C, Wang F, Yuan J. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. MedRxiv Pre-Print. 2020; doi:10.1101/2020.02.11.20021493.

30. National Center for Immunization and Respiratory Diseases Interim Guidelines for Collecting, Handling, and Testing Clinical Specimens from Persons for Coronavirus Disease 2019 (COVID-19); 2020. Available from: `https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html`.

31. Ai T, Yang Z, Xia L. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease. Radiology. 2020;2019:1–8. doi:10.14358/PERS.80.2.000.

32. Porte L, Legarraga P, Vollrath V, Aguilera X, Munita JM, Araos R, et al. Evaluation of Novel Antigen-Based Rapid Detection Test for the Diagnosis of SARS-CoV-2 in Respiratory Samples. Available from `https://ssrn.com/abstract=3569871`; doi:10.2139/ssrn.3569871.

33. Pan Y, Li X, Yang G, Fan J, Tang Y, Zhao J, et al. Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. Journal of Infection. 2020;. doi:10.1016/j.jinf.2020.03.051.

34. Cassaniti I, Novazzi F, Giardina F, Salinaro F, Sachs M, Perlini S, et al. Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department. Journal of Medical Virology. 2020; p. 1–4. doi:10.1002/jmv.25800.

35. Döhla M, Boesecke C, Schulte B, Diegmann C, Sib E, Richter E, et al. Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity. Public Health. 2020;182:170–172. doi:10.1016/j.puhe.2020.04.009.

36. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal. The BMJ. 2020;369. doi:10.1136/bmj.m1328.

37. Foundation for Innovative New Diagnostics. SARS-COV-2 Diagnostics: Performance data; 2020. Available from: https://www.finddx.org/COVID-19/dx-data/.

38. Adhanom T. WHO Director-General's opening remarks at the media briefing on COVID-19 - 16 March 2020; 2020. Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19-

39. Patel NV. Why it's too early to start giving out "immunity passports"; 2020. Available from: https://www.technologyreview.com/2020/04/09/998974/immunity-passports-cornavirus-antibody-test-outside/.

40. Proctor K, Sample I, Oltermann P. 'Immunity passports' could speed up return to work after COVID-19; 2020. Available from: https://www.theguardian.com/world/2020/mar/30/immunity-passports-could-speed-up-return-to-work-after-COVID-19.

41. Fernandez E. The COVID-19 Coronavirus Is Now A Pandemic - Can We Ethically Deal With Lockdowns?; 2020. Available from: https://www.forbes.com/sites/fernandezelizabeth/2020/03/13/the-COVID-19-coronavirus-is-now-a-pandemiccan-we-ethically-deal-with-lock

42. Lilico A. We cannot leave our coronavirus exit strategy to the experts; 2020. Available from: https://www.telegraph.co.uk/politics/2020/04/09/cannot-leave-lockdown-exit-strategy-experts/.

43. Riley-Smith B, Knapton S. Boris Johnson and Donald Trump talk up potential 'game-changer' scientific advances on coronavirus; 2020. Available from: https://www.telegraph.co.uk/news/2020/03/20/boris-johnson-donald-trump-talk-potential-game-changer-scientific/.

44. Cronvall G, Connell N, Kobokovich A, West R, Warmbrod KL, Shearer MP, et al. Developing a National Strategy for Serology ( Antibody Testing ) in the United States. Johns Hopkins University; 2020.

45. Gigerenzer G. Calculated risks: How to know when numbers deceive you. New York, NY, USA: Simon and Schuster; 2002.

46. Beaumont P, Boffey D, Jones S. Singapore coronavirus surge raises fears of post-lockdown breakouts; 2020. Available from: https://www.theguardian.com/world/2020/apr/08/singapore-coronavirus-increase-revives-fears-of-post-lockdown-surge.

47. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review. 2000;42(4):599–653.

## S1: SIR Models

SIR models offer one approach to explore infection dynamics, and the prevalence of a communicable disease. In the generic SIR model, there are $S$ people susceptible to the

illness, $I$ people infected, and $R$ people who are recovered with immunity. The infected people are able to infect susceptible people at rate $\beta$ and they recover from the disease at rate $\gamma$ [47]. Once infected persons have recovered from the disease they are unable to become infected again or infect others. This may be because they now have immunity to the disease or because they have unfortunately died. Figure 8 shows a schematic of the generic model formulation, and how people move between the states. Figure 9 demonstrates the typical disease dynamics, the *Infected* corresponding to the now well known curve that we are trying to flatten.

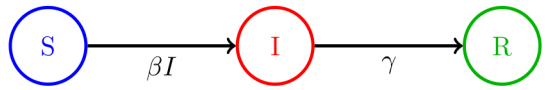


**Fig 8.** Diagram for a basic SIR model. The black arrows show how people move between the different states and the red arrow shows how people become infected.

The SIR model has two ways in which the number of new infections falls to zero. Either the number of susceptible people reduces to a point at which the disease can no longer propagate, perhaps because of a vaccine or natural immunity, or the epidemic stops if the basic reproduction rate of the disease falls below 1 due to social distancing or effective viral suppression.

$$R_0 = \frac{\beta}{\gamma}, \tag{5}$$

.

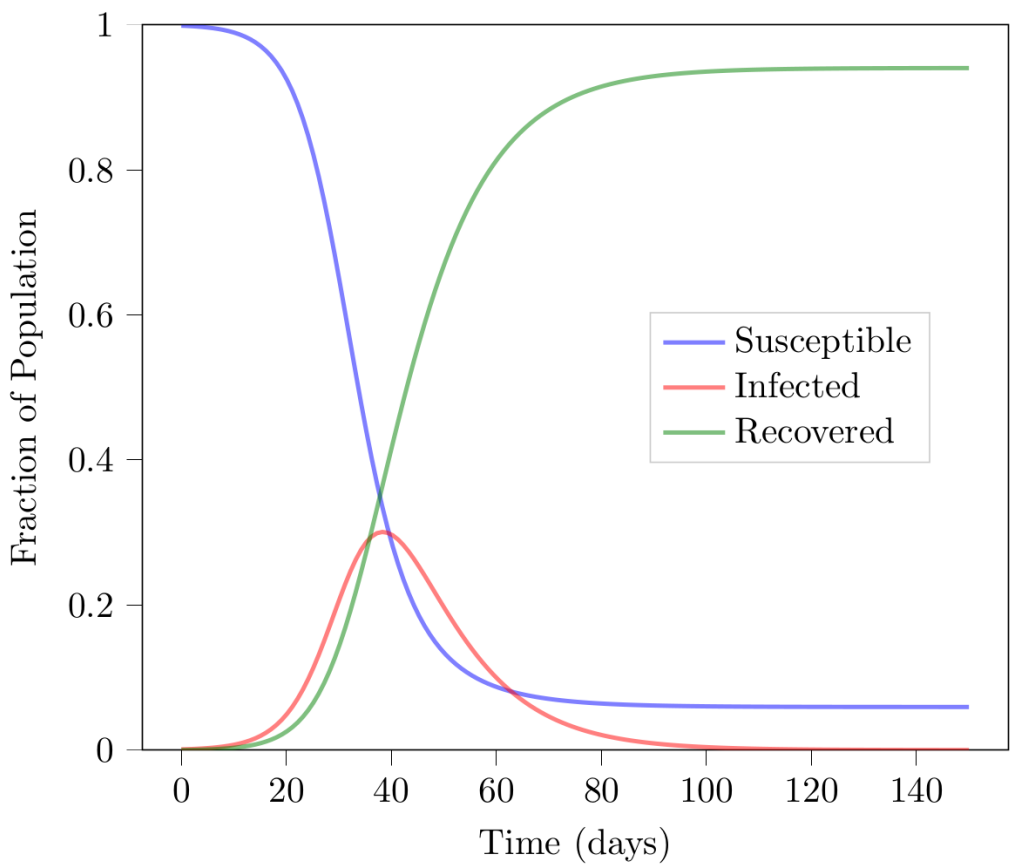


**Fig 9.** Generic SIR model run with $I(t = 0) = 0.0001$, $\beta = 0.3$ and $\gamma = 0.1$.

## S2: Binomial SIR model

The SIR model used in this paper uses discrete-time binomial sampling for calculating movements of individuals between states. For a defined testing strategy, with an active viral test having sensitivity, specificity and capacity of $\sigma_A, \tau_A$ and $C_A$ respectively, an antibody test with sensitivity, specificity and capacity $\sigma_B, \tau_B$ and $C_B$ respectively and a testing prevalence of $p$, these rates are defined as follows:

$$N_A = min\left(C_A, Bin\left(S+I,\rho\right)\right), \tag{6a}$$
$$N_B = min\left(C_B, Bin\left(Q_S+Q_R,\phi\right)\right), \tag{6b}$$
$$T_I = min\left(N_A p, Ip\right), \tag{6c}$$
$$T_S = min(S, N_A - T_I), \tag{6d}$$
$$A = Bin\left(N_B \frac{Q_S}{Q_S+Q_R}, 1-\tau_B\right), \tag{6e}$$
$$B = Bin\left(N_A - min\left(N_A(1-p), I(1-p)\right), 1-\tau_A\right), \tag{6f}$$
$$C = Bin\left(I, \beta\frac{S}{S+I+R}\right), \tag{6g}$$
$$D = Bin\left(min\left(N_A p, Ip\right), \sigma_A\right), \tag{6h}$$
$$E = Bin\left(I-D, \gamma\right), \tag{6i}$$
$$F = Bin\left(Q_I, \gamma\right), \tag{6j}$$
$$G = Bin\left(N_B \frac{Q_R}{Q_S+Q_R}, \sigma_B\right), \tag{6k}$$
$$\Delta Q_S = B - A, \tag{6l}$$
$$\Delta S = A - B - C, \tag{6m}$$
$$\Delta Q_I = C - D - E, \tag{6n}$$
$$\Delta I = D - F, \tag{6o}$$
$$\Delta Q_R = -G, \tag{6p}$$
$$\Delta R = E + F + G, \tag{6q}$$
$$\tag{6r}$$

At each time step $t$, the model calculates the number of persons moving between each state in the order defined above. The use of a binomial model was prompted by a desire to incorporate both aleatory and epistemic uncertainty in each movement. The current approach does not make use of epistemic uncertainty, fixing the model parameters $\sigma_A, \tau_A, \sigma_B, \tau_B, \phi, \rho, C_A, C_B$ and $p$. A discrete time model was selected to allow for comparisons against available published data detailing recorded cases and recoveries on a day-by-day basis.