Is '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], Alex 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[1,2], and Scott Ferson[1]

[1]Institute for Risk and Uncertainty, University of Liverpool
[2]Wirral & Liverpool University Teaching Hospitals

covid19@riskinstitute.uk
April 2020

1 Introduction

The UK government Covid-19 epidemic management strategy has been heavily influenced by the results obtained from Ferguson et al [1] epidemiological models at Imperial College, London early March. Their analysis of the relative impact of different mitigation and suppression strategies has influenced the current approach, with mandated closure of schools and universities and social distancing of the entire population. The analysis highlighted such suppression strategies invite significant challenges to the relaxation of these measures [1]. Without a careful and staggered cessation of the suppression strategies the risk of a second wave becomes significant (possibly of greater magnitude than the first as the SARS-CoV-2 virus is now endemic in the population).

Whilst the *'end-game'* for the current suspension of normal life is now prominent in public discourse, and much attention has been given to the number of tests being conducted, less consideration has been dedicated to the issues of imperfect testing on disease dynamics. PCR testing is a highly sensitive test, however there is a significant difference between the often quoted analytical sensitivity of these tests and the true practical sensitivity. The latter suffering from a multitude of additional uncertainties. Misguided faith in the performance of these tests could have severe implications.

The notion that testing will be a pillar of whatever approach is employed to relax the current social distancing measures in the UK takes two forms. The public are rapidly becoming aware of viral testing lingo. Most are now aware of the 'have you got it?' tests for detecting active cases, and the 'have you had it?' tests for the presence of antibodies, implying some immunity to Covid-19. What may be less obvious to the masses is that in order for these tests to be effectively employed they need to maximise different test characteristic statistics.

For an approach to end the current social distancing measures based on testing active cases, this implies efforts will be made to detect those with an active viral status, who hence have the propensity for viral shedding and to infect others. This would likely look something like the UK's initial containment phase of the Covid-19 epidemic. Where Public Health England aim to find people who currently have the virus, and then contact trace and remove active carriers from the population. This means these tests would need to maximise the sensitivity of the tests, attempting to maximise how good the test is at telling you that you have the disease. In other words, you don't want to miss people that have the virus

*equal contribution

because they could go on to infect further people. 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 they were uncertain of their disease status.

The second testing approach, seeking to detect the presence of antibodies to identify those that have had the disease would be used in a subtly different strategy. This strategy would likely involve detecting those that 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 safe to be allowed to relax their personal social distancing measures. Staggering this across the population would allow society to incrementally recover without the risk of overwhelming healthcare services. 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. In other words, you don't want to tell people they have had the disease when they haven't because you are placing them at risk of still going on to contract the disease. Again, with a false sense of security this is probably even worse than for people to be uncertain about their viral history.

2 What makes a 'good' test good?

In order to answer this question there are a number of important statistics important in assessing the accuracy of diagnostic testing:

- **Sensitivity** (σ) - Out of those who actually have the disease, that fraction that received a positive result.
- **Specificity** (τ) - Out of these who did not have the disease, the fraction that received a negative result.
- **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?

If we tested N people for Covid-19, of which T_+ test positive and T_- test negative. The total number of infected people, S, out of the sample that have been tested is the number of true positives, a plus the number false negatives c. Similarly those who are not infected, W, would be the number of false positives b plus the number of true negatives d. These results could then be tabulated as shown in Table 1, such a table is known as a confusion matrix.

	Infected	Not Intfected	Total
Tested Positive	a	b	T_+
Tested Negative	c	d	T_-
Total	S	W	N

Table 1: Confusion matrix

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

$$\sigma = \frac{a}{a+c} \tag{1}$$

$$\tau = \frac{d}{b+d}. \tag{2}$$

The PPV and NPV can then be calculated using Bayes' rule:

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

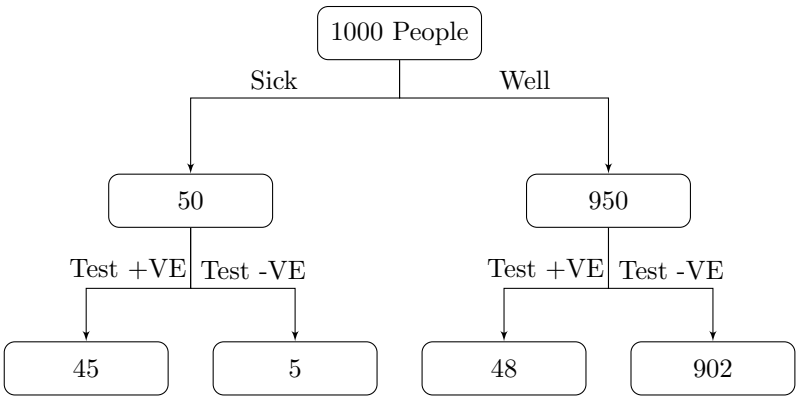

Figure 1: If the prevalence of a disease amongst those being tested is 0.05 then with $s = t = 0.95$ the number of false positives will outnumber the true positives, resulting in a low PPV of around 0.48.

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

Where p is the prevalence of the disease.

In order to improve the diagnostic performance of tests, they are often performed multiple times in order to increase the value of the PPV. In order to do this though, the assumption needs to be made that the two tests are independent of each other, that the result of one test does not impact the result of the other. This assumption could be questionable, for instance if samples are analysed at the same time, there may be systematic biases that mean that this assumption cannot be relied upon.

All of the above statistics require a definitive way of determining the true viral status of a patient, a so-called *gold standard*. This can be very challenging to asses in the midst of such a fast moving epidemic. 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. Not to mention the variation in: patient sampling quality, sample preservation and storage, and sample registration and provenance. Assessing the magnitude of these errors on the practicable performance of tests is very difficult in real time. Even point of care tests are not immune to these errors and are often seen as less accurate than laboratory based tests.

2.1 Why mass testing may be ineffectual

The prevalence of the disease matters. The PPV can vary drastically for different populations with different prevalences. The idea that prevalence depends on the population may seem counter-intuitive to some audiences. For example, if we were to select 5 people from a respiratory ward this week from any hospital in the UK, and 5 people from a street outside the building you are currently sat in, what proportion of each population selected do you think have Covid-19?

To illustrate the impact of prevalence on PPV, for a test with $s = t = 0.95$ if $p = 0.05$ then $PPV \approx 0.48$. 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%. Whereas, using the same tests on a sample with a higher prevalence $p = 0.5$ we find the PPV is higher $PPV = 0.8$, see Figure 2. Similarly, the NPV is lower when the prevalence is higher.

Due to the impact of the uncertainty in testing, it may be the cases that the reported number of infections may be very different to the actual amount of infections even amongst the population that have been tested. This is different to the uncertainty that is caused by not testing even the whole population that is symptomatic.

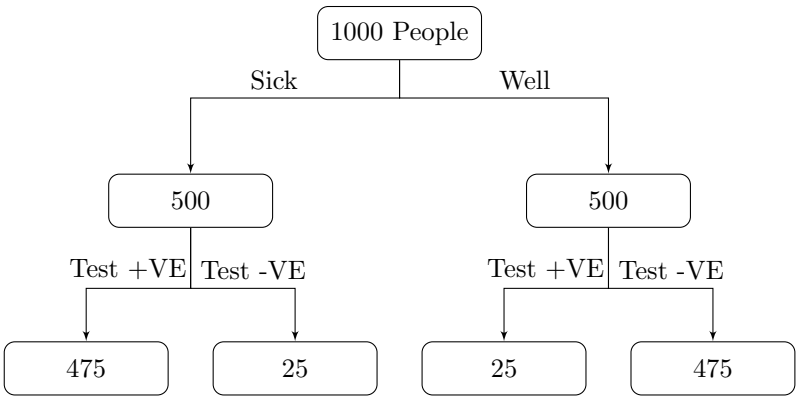

Figure 2: If the prevalence of a disease amongst those being tested is 0.50 then with $s = t = 0.95$ the number of false positives will outnumber the number of true positives, resulting in a high PPV of 0.95.

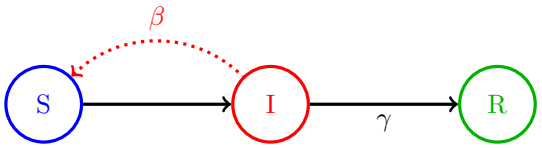

Figure 3: Diagram for a basic SIR model. The black arrows show how people move between the different classes and the red arrow shows how people become infected.

3 So, is no test better than a bad test? or is no test better than any at all?

The UK government have promulgated a snappy phrase to simplify the challenges they have had with the roll-out mass testing. The assertion that 'no test is better than a bad test' may actually be negatively true, no test may also be better than a *'good test'*. It's difficult to know what is (in the UK governments assessment) a good test, or a bad test. That being said, even extremely *'good'* tests may not be good enough.

3.1 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 β and they recover from the disease at rate γ [2]. Once one has recovered from the disease they are unable to become infected again. This may be because they now have immunity to the disease or because they have unfortunately died. Figure 3 shows a schematic of the generic model formulation, and how people move between the states. And, Figure 4 demonstrates the typical disease dynamics, an image that many of the public will now be familiar with, the *Infected* corresponding to the now well known *'curve'* that we are trying to *'flatten'*.

The SIR model implies 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. Alternatively, the epidemic stops if the basic reproduction rate of the disease falls below 1.

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

.

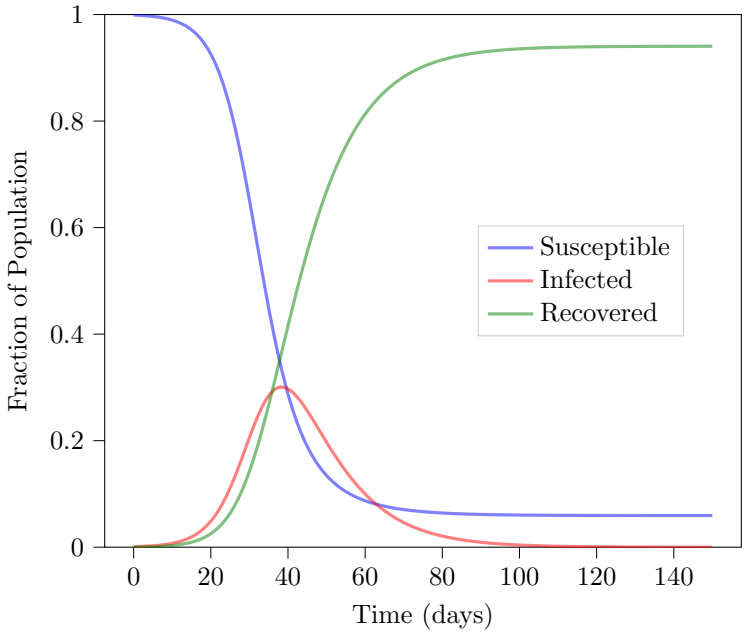

Figure 4: Generic SIR model run with $I(t=0) = 0.0001$, $\beta = 0.3$ and $\gamma = 0.1$.

3.2 SIR model with testing

In order 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 section 3.1 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 finally there is a state for people who have had the disease but are in quarantine, Q_R.

The model evaluations begin with a majority of the population in the Q_S (quarantined but susceptible) state. Analogous to the current state of social distancing measures. Whilst in this initial state the transmission rate of the disease is heavily suppressed. The model evaluates each days average population level dynamic state transitions. Each day there are two possible tests that can be performed:

- An 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. It has a sensitivity of σ_A and a specificity of τ_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 to test whether they have had the disease or not. This test has a sensitivity of σ_B and a specificity of τ_B.

These two tests are used on a proportion of both the susceptible and infected people who are eligible for testing each day, ρ and ϕ respectively. If an infected person receives a (true) positive result then they transition into a quarantine infected state, Q_I, where they are unable to infect anyone else. If an individual in the quarantine recovered state tests (true) positive in an antibody (serology) test then they leave the quarantine state and enter the recovered state.

For this parameterisation the impact of being in the susceptible quarantined state, Q_S, makes an individual entirely 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 and no quarantine is entirely effective, but for the sake of exploring the impact of testing uncertainty these effects are neglected

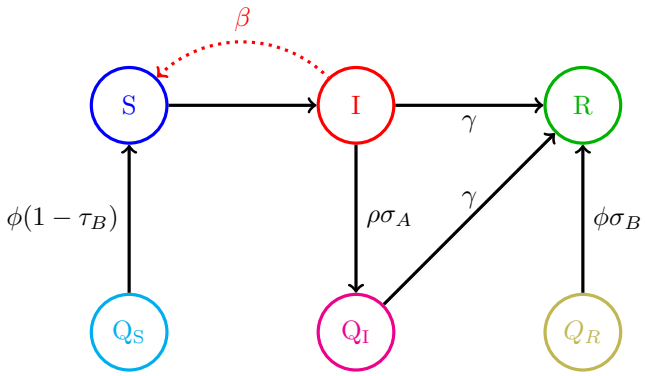

Figure 5: SIR Model used to simulated the effect of mass testing to leave quarantine.

from the model. The participation in infection propagation of individuals in either quarantine state are idiosyncratic, 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 can join the population unlikely to contract the disease again.

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.

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

In order to test the possible impact of testing strategies on the relaxation of current social distancing measures several scenarios have been explored. These scenarios are illustrative of the type of impact, and the likely efficacy of a range of different testing configurations. These are not predictions of what would happen. Each scenario is purely instructive, whatever strategy is employed will not be employed in isolation. It is likely testing will form one pillar of a more elaborate government policy.

- **Immediate end to social distancing scenario**: This baseline scenario is characterised by a sudden relaxation of the current social distancing measures. In practicality this scenario bears no resemblance to the policy that will likely be employed in the UK. It is however a useful initial analysis for bench-marking and comparisons against the other scenarios.

- **Immunity passports style scenario**: A policy that has been discussed in the media, with both support and contention amongst the academic community has been the idea of *Immunity passports*. 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. Whilst the authors are sensitive to the sociological and ethical concerns such an approach engenders, the analysis presented is purely on the question of efficacy.

- **Incremental relaxation scenario**: A phased relaxation of the government 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. Again

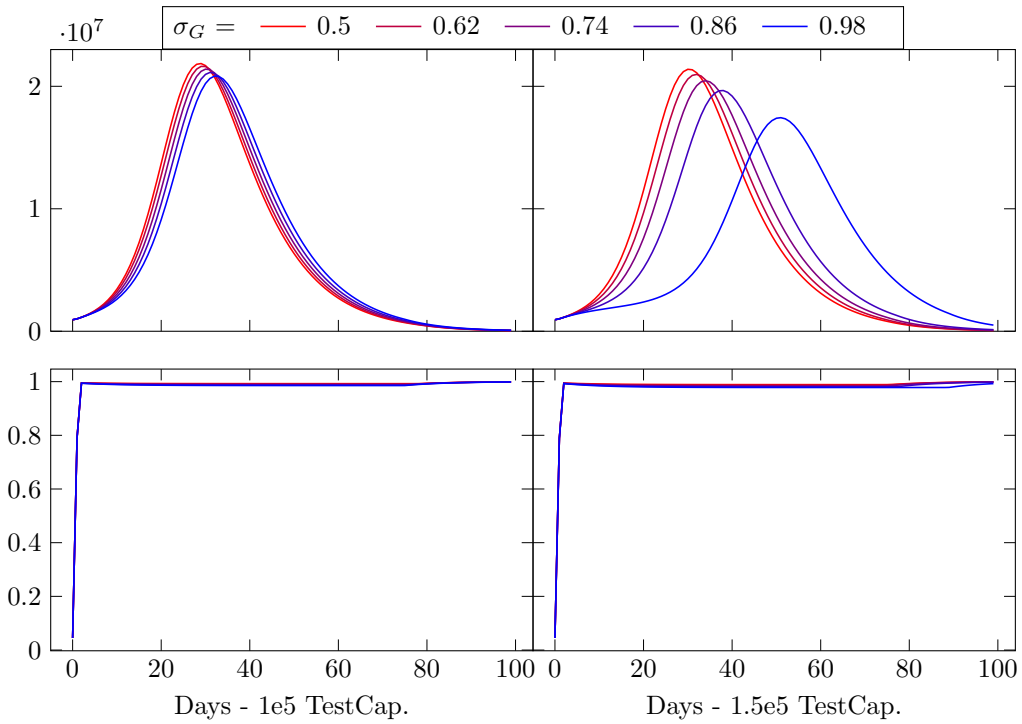

Figure 6: A comparison of different infection test sensitivities σ_G shown from red to blue. Two different infection test capacities are considered. Left: test capacity $= 1 \times 10^5$. Right: test capacity $= 1.5 \times 10^5$. Top: The number of infected individuals ($I + Q_I$ population) over 100 days. Bottom: The proportion of the population which has been released from quarantine ($S + I + R$ population) over 100 days.

the authors make no comment on the sociological and ethical implications of such a strategy, only the efficacy of the approach.

4.1 Immediate end to social distancing 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. Infection testing under our model formulation governs the transition from state I, infected, to state Q_I, infected and in quarantine. Under the same formulation the initial action in the model characterising this scenario is that the majority of the population transition from Q_S to S in the first iteration.

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. 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.

Two evaluations have been conducted. The first using the stated government goal of 100,000 tests per day. It remains unclear whether this aim is feasible, or if this testing capacity would include both forms of tests (antibody and infection). The second evaluation looks at a very optimistic case where we could conduct as many as 250,000 tests per day. 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

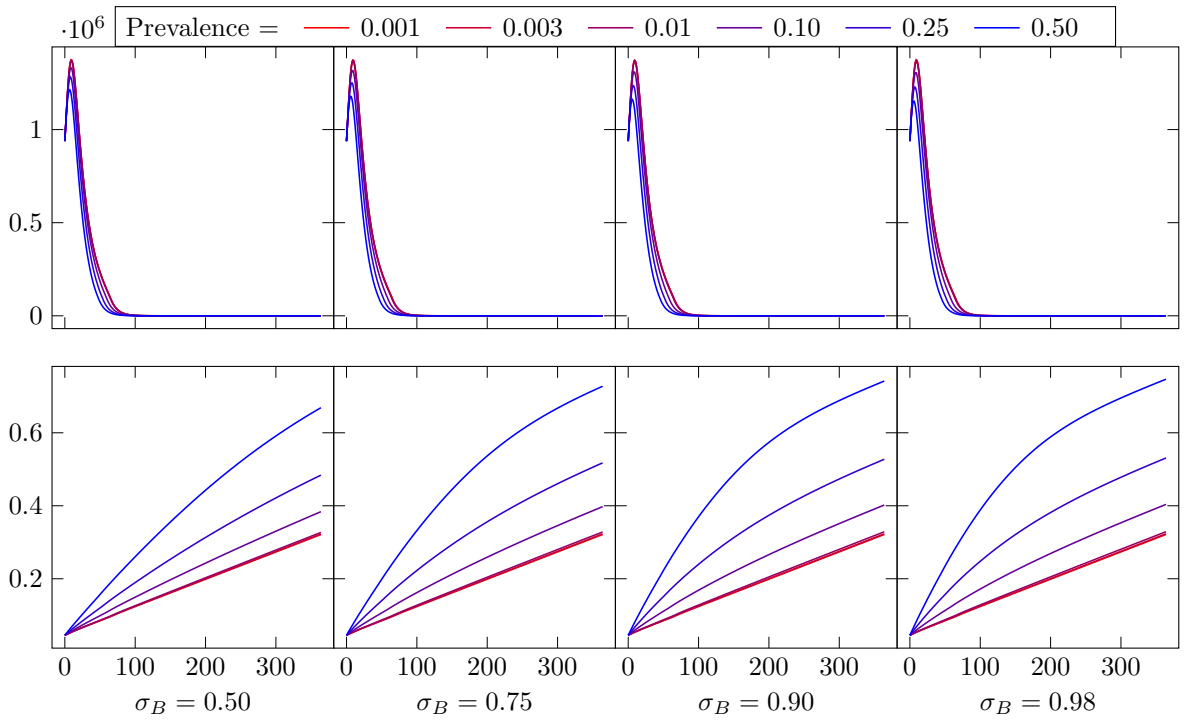

Figure 7: A comparison of different antibody test sensitivities σ_B shown from left to right, with varying levels of prevalence shown from red to blue. Top: The number of infected individuals ($I+Q_I$ population) over one year. Bottom: The proportion of the population which has been released from quarantine ($S+I+R$ population) over one year.

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.

4.2 Immunity passports style 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 referred to as a *'game-changer'*, now exactly what *'game'* it would *'change'* is not clear. That's to say, it's not exactly clear what roll, if any, antibody testing would have in the management of the UK epidemic. It is however fair to say at least some commentators believe that this could have a significant impact on relaxation of social distancing.

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. As was alluded to previously, 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.

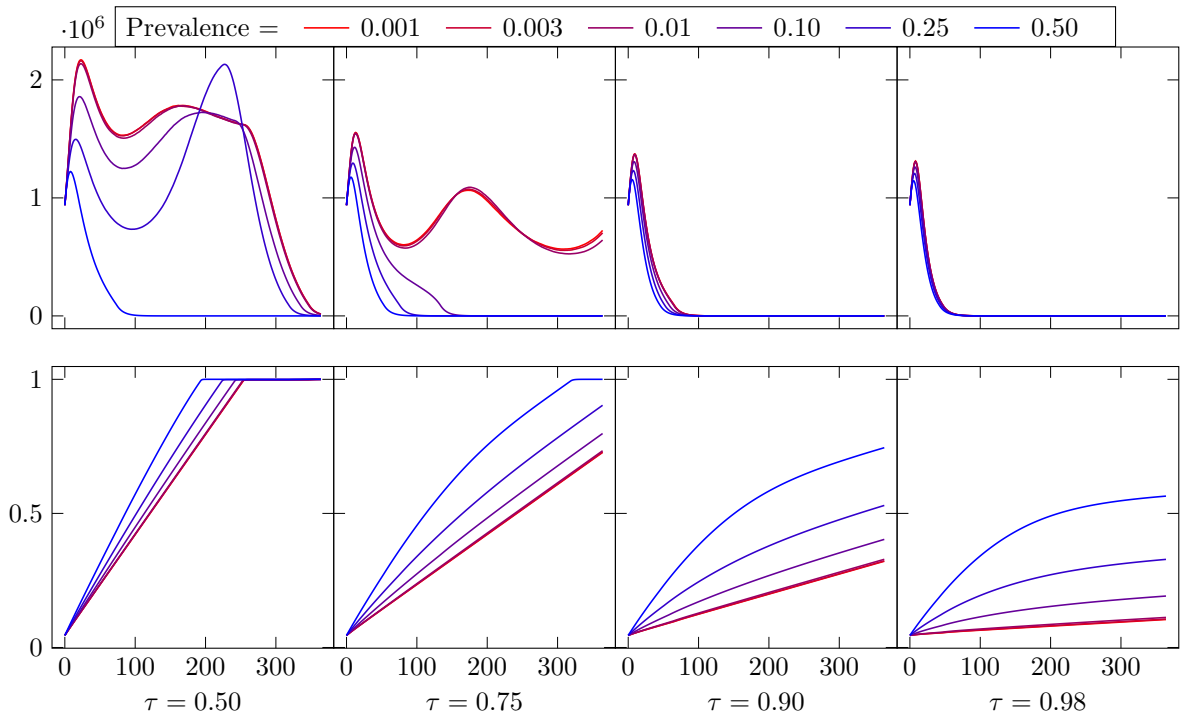

Figure 8: A comparison of different antibody test specificities τ_B shown from left to right, with varying levels of prevalence shown from red to blue. Top: The number of infected individuals ($I+Q_I$ population) over one year. Bottom: The proportion of the population which has been released from quarantine ($S+I+R$ population) over one year.

There are two dominant mechanisms at play here. The obvious is, 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, and as such more people would be correctly (under the paradigms of an immunity passport) allowed to leave quarantine. The slightly less apparent factor is that the lower the prevalence the more people will test positive falsely. (Refer back to section 2.1 for a more complete description of how the prevalence impacts the PPV.) This means more people will relax social distancing measures with a false sense of security, wrongly believing they have some level of immunity when they have not got any. Whilst on an individual basis, and even at the population level, the implications of the behavioural differences could be significant. The model parametrisation here does not include any of these effects, we demonstrate the impact of people reentering the susceptible population assuming their propensity to contract the infection is invariant from those without this false sense of security.

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. 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 class 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 7 and 8 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 really 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 7 as captioned underneath. The specificity for each test in these evaluation was fixed to 90%, which, in our view, would be a reasonable aspiration. Each column corresponds to a different antibody test specificity in figure 8 as captioned underneath. The sensitivity for each test in these evaluation was fixed to 95%, which, in our view, would be an ambitious aspiration, but within the realms of possibility for other rapid antibody ELISA tests. For all model evaluations in figures 7 and 8 the authors modelled a continuing and constant ability to conduct targeted infection 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 prevalence of antibodies in the population. To be clear, this is the proportion of the population that has contracted the SARS-CoV-2 virus and recovered but are in some form of quarantine. Sir Patrick Vallance, the UK Government Chief Scientific advisor in the daily press briefing on the 9th of April stated his belief that this prevalence is likely to be less than 10%, possibly much less. The analysis has explored a range of prevalence from 0.1% to 50%. Figure 7 explores the impact of a variation in sensitivity, from a test with 50% sensitivity (i.e no diagnostic value at all) to tests with the highest feasible 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 in section 2.1, even a highly accurate test produces a very large number of false positives. 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.

It can also be observed in the bottom row of figure 7, the proportion of the entire population leaving quarantine over a year of employing this policy that at low prevalence there is no benefit at all to higher performance 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 that are immune, and gets them back into the community much faster. However this is not the case currently in the UK since, as Sir Patrick stated, the prevalence of antibodies is likely to

be very low.

A more concerning story can be seen when considering the graphs in figure 8. Similar to the analysis previously discussed, now we consider a range of antibody test specificities. Going from 50% (no value in ruling people out) to 98% where false positives are less frequent. It can be observed that 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 in the number of infections. This is because the infection testing would still be employed along side the antibody testing. As the initial wave of people who are given false positive tests for antibodies leave quarantine, this leads to a sharp rise in the number of infections. This would be accompanied by a greater targeting of the infection testing on that population (the infection testing is more effective at detecting more cases because the prevalence of the active virus in the population has increased) initially subduing the rise. This however would be followed by additional waves as further false positive tests for antibodies are observed as the balance is disturbed by changes in the prevalence of antibodies in the quarantined population.

When we consider the bottom half of figure 8 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, and 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. Again it must be stressed, this analysis is not a prediction. Its doubtful that the government policy would be to employ antibody tests in isolation, they would be part of a broader range of measures. However, decision makers and the public need to be aware of the trade-off being made. Professor Gerd Gigerenzer has spent a career trying to educate the medical profession and the public on reckoning with uncertainty, in particular on the dangers of neglecting uncertainties in medical diagnostic testing. In fact, his discussion of AIDS is not too dissimilar to the situation we find ourselves in, particularly if *'immunity passports'* become prominent in the strategy to end the current social distancing measures [3].

4.3 Incremental relaxation scenario

At this point, in the view of the authors, 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, or any other you could imagine. Under the parameterisation chosen for this analysis the model is not sensitive to any particular policy change. The authors make no comment on the impact of any particular modification to the current advice, we simply consider a variety of phased relaxations to the current quarantine. To allow this in the model we consider a weekly incremental transition rate from the Q_S state to the S state, and Q_R state to R. In figure 9, three weekly transition rates have been applied 1%, 5% and 10% of the quarantined population transitioning each week for the duration of the simulation. Whilst in practice this 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 effect of testing capacity and release rate.

The model simulates these rates of transition for a year, and we apply a sensitivity and specificity of 90% for tests. The specifics of each run and detailed in table ¡¡¿¿¿¿¿¡¡¡¡¡¡¿¿¿¿¿. Figure 9 shows five analyses, with increasing capacity for the infection tests. In each, the 3 incremental transition rates are applied with a range of disease prevalence in the population being tested. The PPV, as discussed in section 2.1, has a greater dependence on the prevalence (at lower prevalence) in the tested population than it does on the sensitivity of the tests, the same is true of the specificity and the NPV.

What some readers may find difficult to rationalise is that higher test capacities actually causes a higher peak of infections for the 5% quarantine release rate, see 3rd row of figure 9 with increasing peaks left to right. Under the parameterisation employed in this analysis this has a feasible, albeit counter-intuitive, explanation. When there is the sharpest rise in 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 thousand 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 in the model to be infected, leading to a higher total peak as there is a greater population of susceptible people. This continues until the model reaches effective herd immunity and the number of infected in the population decays very quickly.

Up to a capacity of testing of 120,000 these outcomes are broadly invariant to prevalence of the disease in the tested population at this 10% release rate. 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. It is worth stating, it may be possible to target testing more effectively, and this does have a significant impact on the ability of the testing to flatten the curve once a prevalence of disease in the tested population reaches in excess of 80%. However, any less well targeted testing than this and there is a significant risk that the testing exacerbates the problem, by prolonging high rates of infections in the population, and allowing more people to join the susceptible population that go on to become infected.

With a release rate from quarantine of 5% per week the same increase in infections that were observed at 10% release rate do not occur. Considering the second row of the figure 9, again left to right with increasing capacity for testing, it can be observed that when prevalence of disease is low in the tested population, (i.e 5% poorly targeted tests) the higher test capacities have a negligible effect on the peak or duration of elevated levels of infection. At this lower release rate (5% per week) more targeted tests do become effective, at least at reducing the duration of the peak. How effective testing could be targeted in such a scenario is a matter of debate. The ability of China, for example, to contact trace through the use of surveillance technology may have been the key to their success so far at relaxing quarantine measures, by enhancing the targeting of testing. Whether this could be replicated is uncertain, and in practical terms the infections that you are quite confident have the disease are less likely to go on to infect other people. It's exactly the asymptomatic, or mildly symptomatic patients that you would want to find through testing and have quarantined. So by design its unlikely the prevalence in the test population should reach even 50%.

The analysis of a quarantine release rate of 1% yields some particularly interesting findings. At the slower release rate, the baseline case with no testing, the first column of the top row of figure 9, shows how an initial very small increase in peak infections would be dampened by the slow release of people into the susceptible population from quarantine. This could lead to a natural level of herd immunity, that would lead to a persistent low level of virus throughout the year, as the slow release of people from quarantine continues. This does however only lead to 40% of the population being able to leave quarantine by the end of the year. Possibly too slow in light of other considerations.

For a capacity of 50,000, with well targeted testing of 30% prevalence in the tested population, its possible to suppress the peak infections significantly. The simulation demonstrates at this test capacity, with a sensitivity and specificity of 90%, and a release rate from quarantine of around 1%, we could reasonably expect to see a decline in new infections to the hundreds within 3 months. At which point decision makers may have confidence that a more significant relaxation of measures is safe.

However, the modelling suggests caution should be exercised with the greater numbers of tests. Because the greater quantity of testing rapidly removes most but not all cases of infection, it inhibits the opportunity for those released from quarantine early in the simulation to acquire immunity (these people in the real world would likely be in the demographics most likely to have mild symptoms). The number of susceptible people eventually increases to such a point where the probability that they interact with an infected individual is sufficiently large that a second growth phase of the virus becomes likely. This condition within the model arises as there are few in the population with immunity, because of the success of the early quarantine, and the suppression of infection levels to below a threshold where testing is effective. The proportion of the active population is dominated by those that are susceptible, creating conditions for the second peak that can be observed in the 4th and 5th column of the first row in figure **??**.

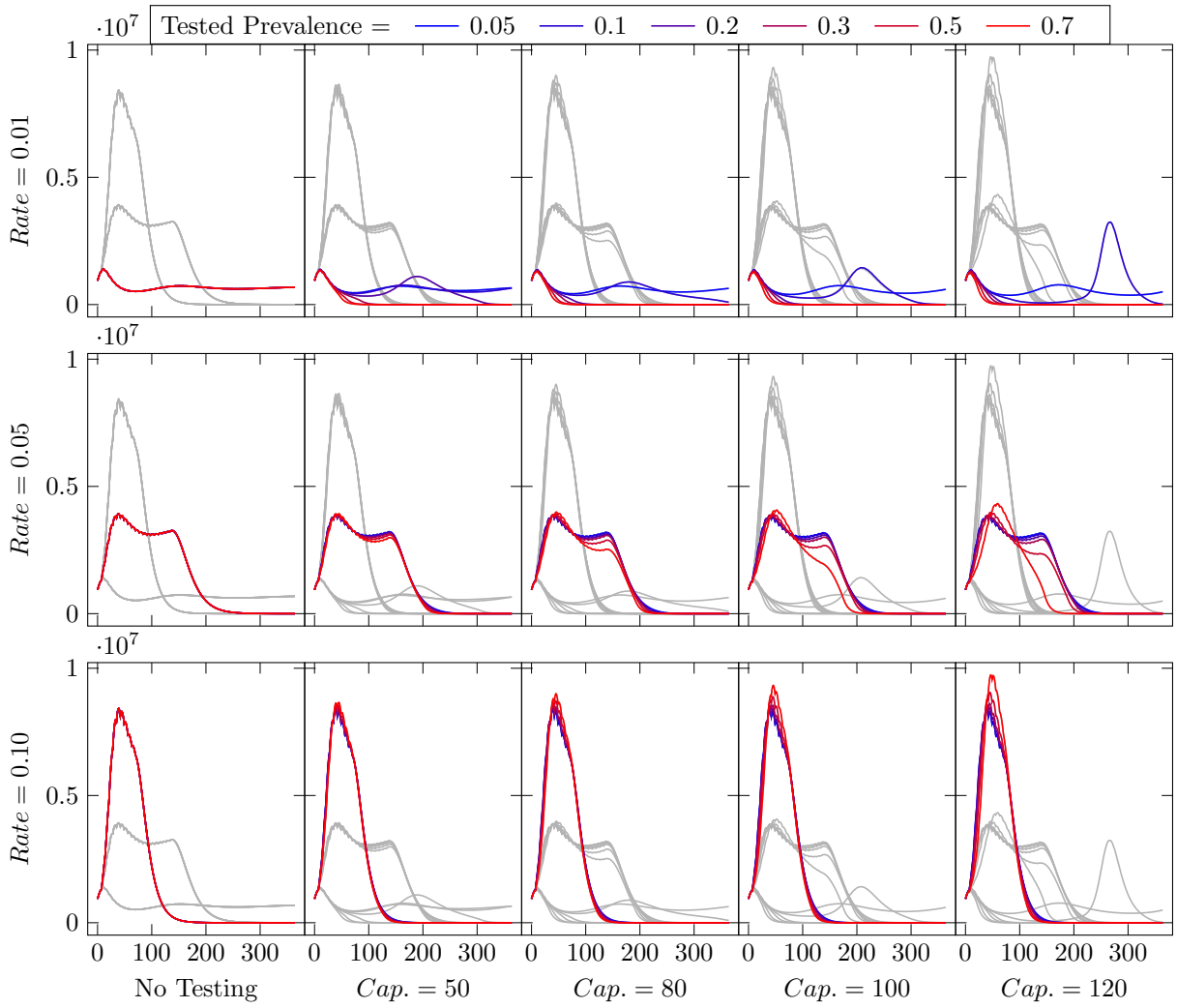

Figure 9: Infection count over 1 year following implementation of a variety of test strategies. Each consider varying: $Rate$, the percentage of the initial quarantined population being released each week; $TestedPrevalence$, the prevalence of the infection in the tested population; and Cap, the infection testing capacity per day.

Model Parameters				Initial Population Split					
σ_G	τ_G	β	γ	Q_S	S	Q_I	I	Q_R	
0.9	0.9	0.32	0.1	0.95	0.034	0.004	0.01	0.001	0
Fixed parameters for the simulations shown in Figure 9									

With a testing capacity of 120,000 tests per day the model suggests a test sensitivity of 95% would be highly effective, and could suppress the growth in infections to a similar level as observed with the lower release rate of just 5%. This demonstrates a very large number of tests, with a very high sensitivity could control the growth rate very effectively. For completeness the model with a test capacity of 120,000 was also evaluated with sensitivities of 80%, 85% and 90%. For the lowest sensitivity we can see that even with this high capacity of testing there is no impact on the peak of infections. However, the two higher sensitivities did allow the testing to pull the peak down by 30%.

When we considered a more modest 10% release rate testing can have a significant effect at a capacity in excess of 80,000. A similar dynamic to that discussed previously explains the increase in peak infections for test capacity less than 100,000. The higher sensitivity tests of this capacity do very significantly suppress the growth. Whilst the lower sensitivity test (80%) has no impact on the peak infections, the higher sensitivity tests are able to suppress the growth completely. A test capacity of 100,000 with a sensitivity of 90% would be sufficient to allow release from quarantine rate of 10% with a negligible increase in peak infections compared to the slower 5% quarantine release rate.

The slowest release rate considered in the analysis of 5% of the population in quarantine released incrementally each week significantly dampens the infection growth rate, and would suppress the peak to 3 million infections. 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 decline faster, reducing the length of necessary wide-scale social distancing.

With a testing sensitivity of 90% or greater, and a test capacity of 100,000 units per day, the analysis demonstrates a very strong ability to subdue the growth in infections almost completely, and thus would have a profound impact on the management of the UK epidemic. If testing capacity could be increased to 120,000 or more this modelling suggests it is feasible the UK could return to some semblance of normality within months, with a relaxation of the most severe social restrictions.

5 Conclusions

This analysis does support the assertion of many government officials that a *'bad test'* is not worse than no tests, but a *'good test'* is only effective in sufficient numbers to get ahead of the virus. This analysis is not a prediction. 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. And certainly no conclusions should be drawn about the specific policy changes that would be necessary to safely end the current social distancing measures.

There are firm conclusions however that this analysis supports:

- Diagnostic uncertainty can have a large effect on the epidemic dynamics of Covid-19 within the UK.
- 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 low, its 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.
- Antibody testing, with a high specificity may be very useful on an individual basis, certainly has scientific value, and could reduce risk for key workers.

- The capacity for infection screening needs to be significantly increased if it is to be used to relax quarantine measures, but only if the sensitivity is very high. A sensitivity of less that 90% would lead to the prolonging of lock-down measures, and anecdotal evidence suggests the current testing is not this accurate.
- It was known when the current lock down measures were enacted, it would be very difficult to end them. Testing at the population level is unlikely to have a significant impact on reducing the length of these measures. At this point the incremental suspension of measures accompanied with highly targeted infection testing is the only feasible way measures can be safely removed.
- One interpretation of these results is that countries that had mass testing regimes right at the beginning of the pandemic but had much lower case fatality rates may have been reporting a large number of false positives. The authors do not wish to imply that this was indeed the cases, only that the results do not exclude the possibility.

A Generic SIR differential Equations

The following set of differential equations govern the behaviour of the basic SIR model:

$$\dot{S} = -\beta SI \tag{6a}$$
$$\dot{I} = \beta SI - \gamma I \tag{6b}$$
$$\dot{R} = \gamma I \tag{6c}$$

Where β is the infections rate of the disease and γ is the recovery rate.

B Novel SIR model with testing

The following set of differential equations are used to describe the dynamics of the SIR model with the effect of imperfect testing regimes:

$$\dot{S} = \phi(1-\tau_B)Q_S - \beta SI \tag{7a}$$
$$\dot{Q_S} = -\phi(1-\tau_B)Q_S \tag{7b}$$
$$\dot{I} = \beta SI - \rho\sigma_A I - (1-\rho\sigma_A)\gamma I \tag{7c}$$
$$\dot{Q_I} = \rho\sigma_A I - \gamma I_+ \tag{7d}$$
$$\dot{R} = (1-\rho)\gamma I + \gamma Q_I. \tag{7e}$$

Where β is the infections rate of the disease and γ is the recovery rate. ρ is the proportion of infectious people that get tested, ϕ is the proportion of

References

[1] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C Okell, Sabine Van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick Gt Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, and Azra C Ghani. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Technical report, Imperial College COVID-19 Response Team, 2020.

[2] Herbert W Hethcote. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4):599–653, 2000.

[3] Gerd Gigerenzer. *Calculated risks: How to know when numbers deceive you.* Simon and Schuster, 2002.