RESEARCH ARTICLE

Human agency and infection rates: Implications for social distancing during epidemics

Christopher Bronk Ramsey *

School of Archaeology, University of Oxford, Oxford, United Kingdom

* christopher.ramsey@arch.ox.ac.uk

Abstract

Social distancing is an important measure in controlling epidemics. This paper presents a simple theoretical model focussed on the implications of the wide range in interaction rates between individuals, both within the workplace and in social settings. The model is based on well-mixed populations and so is not intended for studying geographic spread. The model shows that epidemic growth rate is largely determined by the upper interactivity quantiles of society, implying that the most efficient methods of epidemic control are interaction capping approaches rather than overall reductions in interaction. The theoretical model can also be applied to look at aspects of the dynamics of epidemic progression under various scenarios. The theoretical model suggests that with no intervention herd immunity would be achieved with a lower overall infection rate than if variation in interaction rate is ignored, because by this stage almost all the most interactive members of society would have had the infection; however the overall mortality with such an approach is very high. Scenarios for mitigation and suppression suggest that, by using interactivity capping, it should be possible to control an epidemic without extreme sanctions on the majority of the population if $R0$ of the uncontrolled infection is 2.4. However to control the infection rate to a specific level will always require measures to be switched on and off and for this reason elimination is likely to be a less costly policy in the long run. While social distancing alone can be used for elimination, it would not on its own be an efficient mechanism to prevent reinfection. The use of robust testing, quarantining, and contact tracing would strengthen any social distancing measures, speed up elimination, and be a better tool for the prevention of infection or reinfection. Because the analysis presented here is theoretical, and not data-driven, it is intended to be a stimulus for further data-collection, particularly on individual interactivity levels, and for more comprehensive modelling which takes account of the type of heterogeneity discussed here. While there are some clear lessons from the simple model presented here, policy makers should have these tested and validated by epidemiological specialists before acting on them.

Introduction

The use of social distancing has become a critical tool for countries attempting to deal with the outbreak of COVID-19 [1]. The difficulty they face is particularly in coming up with strategies which are sustainable in the long-run, or which are capable of eliminating the virus entirely. This paper sets out to explore one aspect of this question with a very simple epidemic progression model focussed on looking at the implications of variability in individual contributions to the spread of infections within an epidemic.

The key parameter for understanding the epidemic dynamics is $R0$, which is the average number of other individuals infected by someone who carries the disease when people are operating normally. If this figure is above 1, then exponential growth is expected until there is saturation due to some form of herd immunity. If the figure is less than 1 then the disease will, in time, be eliminated. As $R0$ is the initial rate, at this point it is assumed that overall infection rates are low and both immunity and the probability of passing on the infection to someone already infected can be ignored.

In many ways, the best approach is to use large-scale epidemic simulation [1, 2] or to use a complex network model [3], but it is worth testing these approaches against simpler models which can be used to explore aspects of the problem. In this paper, a totally theoretical probability-based approach is taken, which treats the population as a whole, and which specifically addresses variation in social interaction of individuals, with the aim of understanding how different human agents contribute to the spread. This model is not useful for studying geographic spread and, because it does not have any regional granularity, it cannot address specific questions of how to deal with households or institutions, though in principle the model could be considered for institutions as approximately independent units, where infection rates are low. The purpose here is not to replace fuller simulation or network based studies but to provide a model that takes account of heterogeneity at the individual level, and which is simple enough for non-specialists to use and to understand why some strategies might be more effective than others. The approach is also one that could easily be incorporated into more complex models which currently assume homogeneity at the individual level.

Methods

A simplified theoretical model

Part of the motivation for this theoretical approach is the limited nature of the social contact data which underly simulation models, and are required for quantifying networks. In particular, the types of survey data used may under-report chance interactions [4], and indeed the scale of variation in the length, intensity and closeness of contacts is hard to elicit from such data. The amount of interaction any individual has in a complex modern society will vary over a very wide scale, and probably over a wider scale than contact survey data would suggest. This is partly due to the upper limit on the number of contacts reported, either because the respondent cannot remember them, or because they do not necessarily think they are significant. For the purposes of this model the number of contacts is, in any case, not as significant as total one-one contact time for virus transmission.

Here likely ranges of one-one contact times are estimated, and a simplified model presented which uses this parameter to model infection spread within the population.

A stepwise approach to epidemic progression

The parameter $R0$, if considered as a global model parameter, is the average expected number of new infections arising from each randomly selected infected individual from an overall

population of N. In a stepwise process, the probability that each individual i is infected in cycle j can be defined as p_{ij}. If an individual is infected in any cycle, the expected number of secondary infections in the following cycle is defined as t_i; this is a measure of how intensely that individual interacts with others both socially and through their work; it can be considered as a personal R factor. From this it is possible to extract the expected number of infections at one iteration of the cycle M_j and the number in the next cycle M_{j+1}.

$$
\begin{aligned}
M_j &= \sum_1^N p_{ij} \\
M_{j+1} &= \sum_1^N p_{ij} t_i \\
R0 &= M_{j+1}/M_j
\end{aligned}
\qquad (1)
$$

Human interaction sensitivity

In considering the effect of individuals within the population, it is important to note that $p_i j$ and t_i are likely to be very highly correlated. This is because a person who has a high probability of passing an infection on is also much more likely to pick it up. A realistic assumption for air-borne viruses would be that these two are actually directly proportional because they both arise from the amount of time spent in close proximity to others. This can be used to derive a value for $R0$ from the average transmissibility as $\bar{t}$:

$$
\begin{aligned}
\bar{t} &= \frac{1}{N}\sum_1^N t_i \\
p_{ij} &= \frac{t_i M_j}{N\bar{t}} \\
M_{j+1} &= \frac{1}{N\bar{t}}\sum_1^N t_i^2 M_j \\
R0 &= \frac{1}{N\bar{t}}\sum_1^N t_i^2
\end{aligned}
\qquad (2)
$$

This simple model, with its non-linear dependency on t_i^2, suggests that the $R0$ parameter for the population will be dominated by the contribution from those with high t values.

Variability in interactivity

The parameter t for each individual is likely to be very variable within a population and, the exact numbers will vary between countries and regions. There are some good data on contact numbers and intensity [4, 5], but these are not in exactly the form needed for this model. However, some reasonable upper and lower limits can be estimated. At the upper limit will be those who spend most of their day in close proximity to four other individuals in crowded spaces (*ca.* 50 person hrs/day), and at the other end of the scale, people who only have close interaction very rarely when doing essential errands, or those living alone who are rarely visited (*ca.* 0.5 person hrs/day). Whether contact within the household is considered is a difficult question, because households are particularly integrated. It might be better to consider each household as a super-individual—so with extra interactions and greater chances of picking up the infection together; this issue will not be addressed here, but obviously should be within more complex simulation studies. The exact numbers are not important but the variability is, and here it

is assumed that there is likely to be a 95% probability range of about two orders of magnitude in t. The distribution $f(t)$ will be estimated to be log normal (using the estimates above, the median would be *ca*. 5 person hrs/day).

$$f(t) = \text{Lognormal}(\mu, \sigma^2) \tag{3}$$

The range in interactivity is broadly in line with the findings in Kissler [5] where mobile phone data are used to track interactions; in that study, whatever minimum range you set for an interaction (1, 2 or 3m), the 97.5th centile of interactivity is just over 20 times the median (compared to 10 times in the parameters chosen here); the absolute interaction times are lower in that study but it does not cover all types of interaction, and in any case the absolute value is not important in this paper.

Integrating over the whole population gives:

$$\begin{aligned} \bar{t} &= \int t f(t) dt \\ &= e^{\mu + \sigma^2/2} \end{aligned} \tag{4}$$

$$\begin{aligned} R0 &= \frac{1}{\bar{t}} \int t^2 f(t) dt \\ &= \frac{1}{\bar{t}} e^{2\mu + 2\sigma^2} \\ &= e^{\mu + 3\sigma^2/2} \\ &= \bar{t} e^{\sigma^2} \end{aligned} \tag{5}$$

This leads to the expected result that, if all individuals in the population have the same t value ($\sigma = 0$), this will also be $R0$ for the population. However, if the 95% probability range for t covers a range α then $\sigma = \ln\alpha/4$, implying that with $\alpha = 100$ as estimated above, $R0 = 3.8\bar{t}$. While this is numerically correct, the very strong weighting to high t values makes this highly dependent on the upper limit of interactivity, and because there are practical limits to the interactions any individual can have, we expect the upper tail of the distribution to be attenuated; if the upper value of t is capped at the 2σ level this is reduced to $R0 = 2.4\bar{t}$ which may be more realistic, and provides a more conservative estimate of the factor. For comparison if α is only 10 then $R0 = 1.4\bar{t}$, whether or not a cap for a reasonable upper limit is used.

Dynamic modelling

The above framework has been developed into a simple dynamic step-wise model. The population is divided into 800 quantiles, each with a different t value, covering $\pm 4\sigma$. Initially these are seeded with a probability of infection set for the particular model run. Using the t values for each quantile the expected number of potential infections can be calculated directly for the population as a whole. These are then partitioned to the population in proportion to the t values of all individuals. The actual infection probability for each quantile is then reduced by the proportion of that quantile which has already been infected.

Unless otherwise stated an $R0$ of 2.4 will be used for comparison with the estimate for COVID-19 in Ferguson [1], with a 95% range of interactivity of 100 and a maximum interaction rate set at 10 time the median. This gives a $\bar{t}$ of 0.995 and a median value of t of 0.558. All outputs are expressed in cycles of infection, following the logic of this simple model; in the case of COVID-19 this is expected to correspond roughly to weeks. In practice with a distributed population, everything would be expected to lag more than this model would suggest. It is

also important to note that the parameter modelled here is new infections. Symptoms will lag this and detected cases (especially if they are based on hospitalisation date) will lag further, and deaths further still. This model is best suited to looking at endemic infection within a single integrated population. The initial seed infection rate probability is chosen to be 0.0001 which is the level at which it is very obvious there is an issue; the seed infection is assumed to be distributed in proportion to individuals' interactivity which fits with the overall logic of this model.

For many of the scenarios it is useful to split the population into groups with different interventions. Where applicable in this paper three groups have been used: the majority (60%) of the population is in group 1 with the remainder split between the vulnerable (20%) in group 2, and key workers (20%) in group 3. In general group 1 will have extra measures for shielding and group 3 will have fewer restrictions so that they can perform their key tasks. Where mortality is discussed, this is assumed to be 0.5% for groups 1 and 3, and 5% for group 2; this gives an infection fatality ratio (IFR) of 1.4% overall which is the upper limit of what is expected [6].

For scenario testing, threshold levels and a response delay can be set in the model. The levels which have been chosen for the scenarios below (where relevant) are 0.0002 (double the seed probability), 0.0004 and 0.001. The response time is 3 cycles which is probably the minimum practical, given that cases have to be detected and some notice of measures has to be given. Longer response times have fairly predicable effects, particularly in increasing the case load in the initial pulse.

In general four different levels of response using two different strategies have been considered. The first strategy, based on the analysis above is to limit the interactivity of the most interactive individuals, essentially capping this at a particular point. The second strategy is to reduce everyone's activity by the same factor. In practice most policy interventions are likely to be a mixture of these. The four levels of interaction considered are:

1. Limited distancing: this entails a capping of activity at the $+1\sigma$ level, which with the parameters above would entail $\sqrt{10}$ above the median level of interaction (*ca* 15 person hrs/day). An approximately similar effect can come from reducing the whole population's activity by a factor of 0.5.

2. Enhanced distancing: this entails a capping of activity at the $+0.5\sigma$ level, which with the parameters above would entail $\sqrt[4]{10}$ above the median level of interaction (*ca* 9 person hrs/day). An approximately similar effect can come from reducing the whole population's activity by a factor of 0.33.

3. Strong distancing: this entails a capping of activity at the median level, which with the parameters above would entail at the median level of interaction (*ca* 5 person hrs/day). An approximately similar effect can come from reducing the whole population's activity by a factor of 0.2.

4. Isolation: this entails a capping of activity at the -1σ level, which with the parameters above would entail $\sqrt{10}$ below the median level of interaction (*ca* 1-2 person hrs/day). It is hard to replicate this with population-wide reductions in activity; here a factor of 0.1 is used (the estimated ratio required is about 0.07).

These levels are chosen specifically because limited distancing results in an R_{eff} just above 1 and enhanced distancing results in an R_{eff} of just below 1, which allows for control of the epidemic progression. It is helpful to consider caps in terms of natural variation in activity since it should be possible for people to reduce their interaction levels to the median level, since half the population does this anyway. It is much harder to assess whether everyone reducing their interaction level by a factor of over 2 is really practical.

Results and discussion

Implications of the theoretical model framework

The analysis above shows the dominant effect of the upper end of the t distribution on $R0$ and implies that an effective way to reduce the growth rate to some lower effective value (R_{eff}) is to reduce σ rather than reducing $\bar{t}$, or more specifically to reduce the interactivity of those members of the population with the highest values of t.

Overall reductions in interaction. One way to reduce R_{eff} is to reduce the interaction level of the entire population by the same factor. It can be seen from Eq 5 that this simply requires a reduction in $\bar{t}$ of the $R_{eff}/R0$. For example to reduce an $R0$ of 2.4 to $R_{eff} < 1$, requires a drop in interactions of a factor of more than 2.4. This is probably very difficult for those with lower interaction levels to achieve and is almost certainly not the most efficient way to do this.

A more practical approach to overall reductions in effective interaction rate is the widespread use of masks, the practice of maintaining distance (typically 1.5 or 2m) between individuals, and limits on group sizes. All of these have the advantage of reducing viral transmission while allowing for interaction in work or social settings. However, such approaches are not universally adopted and so while the median t and $\bar{t}$ might be significantly reduced it is likely that the range of variation in t_i values will if anything be even higher. On their own these may not be effective, especially if some of most interactive individuals do not adopt these measures.

Partitioning of interaction. Perhaps the most interesting impact of the high correlation between p_{ij} and t_i is the effect on the partitioning of interaction between individuals. This can also be seen directly from Eq 5, because a more even partitioning of interactions between individuals will have a significant effect on R_{eff}. In the limit that all interactions are evenly distributed through the population, it might be possible to reduce R_{eff} by the factor of 2.4 while keeping the overall interaction levels ($\bar{t}$) the same. This is not going to be practical but it underlines the important of considering partitioning in designing distancing strategies.

This point can be illustrated by a couple of thought experiments relating to people who might have high interaction rates.

In the first case we consider a workplace where there is just one particular task which involves a lot of interaction with others; this could be for example a receptionist or sales representative. The employer has the choice of whether to split this job between two people or to give it to just one. If the job is split then each of those involved will have only half the probability of infection in this role and also only half the probability of passing it on. For the population as a whole the chance that this particular task resulting in infections being passed on is reduced by a factor of two simply by splitting the job between two people. Of course in a workplace a much better solution might be to allow for proper social distancing and personal protective equipment (PPE), to reduce the risks until they are insignificant.

Much more difficult to address in regulation is social activity between friends outside the workplace. Consider the case of five friends, part of a much larger network, who are normally in regular social contact for long periods, several times a week. Given the need for social distancing, all five individuals might decide to limit their social activities to one evening a week, otherwise remaining at home. Supposing instead of this, four of them decided to stay at home altogether, and just one of them spent five evenings out a week socialising. On the face of this the amount of activity is the same (measured for example by footfall in social venues, or travel). However that one individual has five times the probability of picking up an infection as in the first option, and also five times the chance of passing the infection on; their overall impact on the propagation of the epidemic is 25 times that of an individual going out once a week. Even

taking account of the fact there is only one such individual this option is five times worse than with an even division of socialisation time.

Given that social interaction is likely to be the most difficult aspect to regulate, individual human agency becomes critical to the control of epidemics.

Proportions of the population contribution to $R0$. Another way to look at this is to consider the effect of different sections of the $f(t)$ distribution on $R0$. If the log normal distribution truncated at $+2\sigma$ is used, as discussed above, 50% of the value of $R0$ is generated by the most interactive 5% of the population. Likewise the bottom 50% of the population in interaction terms only contributes about 2% to the total, which makes it clear why modifying their behaviour is unlikely to have significant effect: these are people who are unlikely to get the infection and unlikely to pass it on.

Targeting of interventions. The most efficient way to implement social distancing measures is to target them at the most active members of the population. Indeed, as seen above in the discussion of partitioning, the key thing is to limit the maximum amount of interaction any individual has. Individuals with very high interaction levels during an epidemic present a risk both to themselves and to the population as a whole. This is well known from anecdotal discussion of super-spreaders, but is numerically clear from Eq 2.

This is exactly what you would expect from network analysis [3], where immunisation of high degree nodes is shown to be far more effective than random immunisation. Essentially this is the same strategy, except here it is proposed that reduced interaction of highly interactive individuals is used where immunisation is not available.

Conversely the section of the population which is least interactive has almost no influence on R_{eff} at all. It is clear from this analysis, if not already, that trying to drive down interaction levels that are already far below the median will have almost no effect on R_{eff}. However, any reduction in interaction will reduce the risk of infection for that individual and so may still be a useful measure, especially for those at greatest risk from the infection.

Putting some numbers on this from the estimates above, if the maximum interaction any individual can have is limited to 1σ above the median level in the log normal distribution this will reduce R_{eff} by a factor of 0.51, so almost enough on its own to reduce the an $R0$ of 2.4 to an R_{eff} of 1. If the median contact was 5 person hrs/day this would imply an upper limit of 15.8 person hrs/day. A limit set at the median value would reduce R_{eff} by a factor of 0.20, certainly low enough to eliminate the disease, and somewhere between these two would probably be the optimum (see scenarios below).

These estimates are based on a whole population group, and more granular information could be used to inform p_{ij}, since this will also depend on local infection rates over time: it would make sense to use local case rates rather than national ones. Another important thing to consider is that it is the average interaction rate over a time period of the incubation that is relevant. Temporary high interaction levels, for example at a single event, will be strongly mitigated if combined with the periods before and after have much lower interaction rates. On the other hand, residential events with intense interaction over several days will probably have the greatest impact on raising R_{eff}.

Contact tracing, testing and immunisation. This analysis also has implications for contact tracing [7], because it shows that the people for whom this is most important are those who interact most strongly. Thus if contact tracing can be made to work efficiently on this group alone, it could help to limit the effect of the highly active groups on the overall statistics. Likewise, if it is not possible to reduce interaction rates (for front-line medical staff for example), then the risks of passing infections on can be reduced by other measures such as personal protection and regular testing. It is likely that for the least interactive portion of the population contact tracing, regular testing and immunisation would be less likely to have a major impact

on the progression of the epidemic, but of course may be useful for the protection of the most vulnerable.

In a period when there are still significant risks, one way to allow interaction risk events to proceed, and institutions to remain open, would be to make use of contact tracing and testing to ensure that p_{ij} is low for all individuals at the start.

There are also synergies here with contact tracing strategies, and in particular with the use of mobile phone apps for implementation of this. These apps would be ideal for alerting individuals that their overall interaction rate was high so that they could socially distance for a while to reduce this. They would also be an ideal way of identifying those individuals who would most benefit from immunisation when and if this becomes possible [3].

Scenario modelling

Many of the findings of the scenarios here will be similar under other models; the key here is control of R_{eff}. However, the actual dynamics and mortality rates are model dependent and, as this is a purely theoretical model, should not be treated as predictions of what would happen in practice. Further simulation models should be used to test the implications presented here.

No intervention.   One possible scenario is that there is no intervention. In this case the model runs until herd immunity is reached (see Fig 1). The peak in new cases occurs around cycle 9, with about half the overall cases within three or four cycles. In this model herd immunity is reached with an overall infection rate of 41%. This is lower than the 81% estimated in the simulation model [1], which could be due to limitations of the model here, or to underestimation of interactive variability in the simulation. The herd immunity level is dependent on the variability estimates, but even with a factor of only 10 it is still 73%, and to get to an estimate of 81% it is necessary to lower this to a factor of only 3, which is unrealistic.

The overall mortality rate is 0.0057, assuming that overload in the health systems does not raise the prior assumptions about mortality rate.

The levels required for herd immunity here are due to the treatment of quantiles with different activity levels. In this scenario the very active individuals are almost all infected (Fig 1D) and thereby reduce the overall rate of rise for the population as a whole, until R falls below 1. This is something which needs further investigation within the simulation studies and the data from the epidemic.

Protection.   This is the same as the no intervention scenario but introduces strong distancing for the vulnerable group at a case threshold of 0.0002 and isolation at 0.0004. Under this model the overall progression of the epidemic is similar but the mortality rate drops by a factor of about 2 to 0.0027 (see Fig 2).

Mitigation.   One possible option considered is to delay and flatten the peak, but this turns out to be quite difficult to achieve in this model. It requires a choice of those interventions which reduce R_{eff} to just around 1 which is very hard to do. Here it is achieved by setting the following limits at the three thresholds:

- 0.0002: the vulnerable group goes onto strong distancing while all others go onto limited distancing.

- 0.0004: the vulnerable group goes into isolation.

- 0.001: the main group goes into enhanced distancing.

This is just enough to control the disease at a new case rate of about 0.001 (see Fig 3). However the controls have to be cycled on and off for the main group.
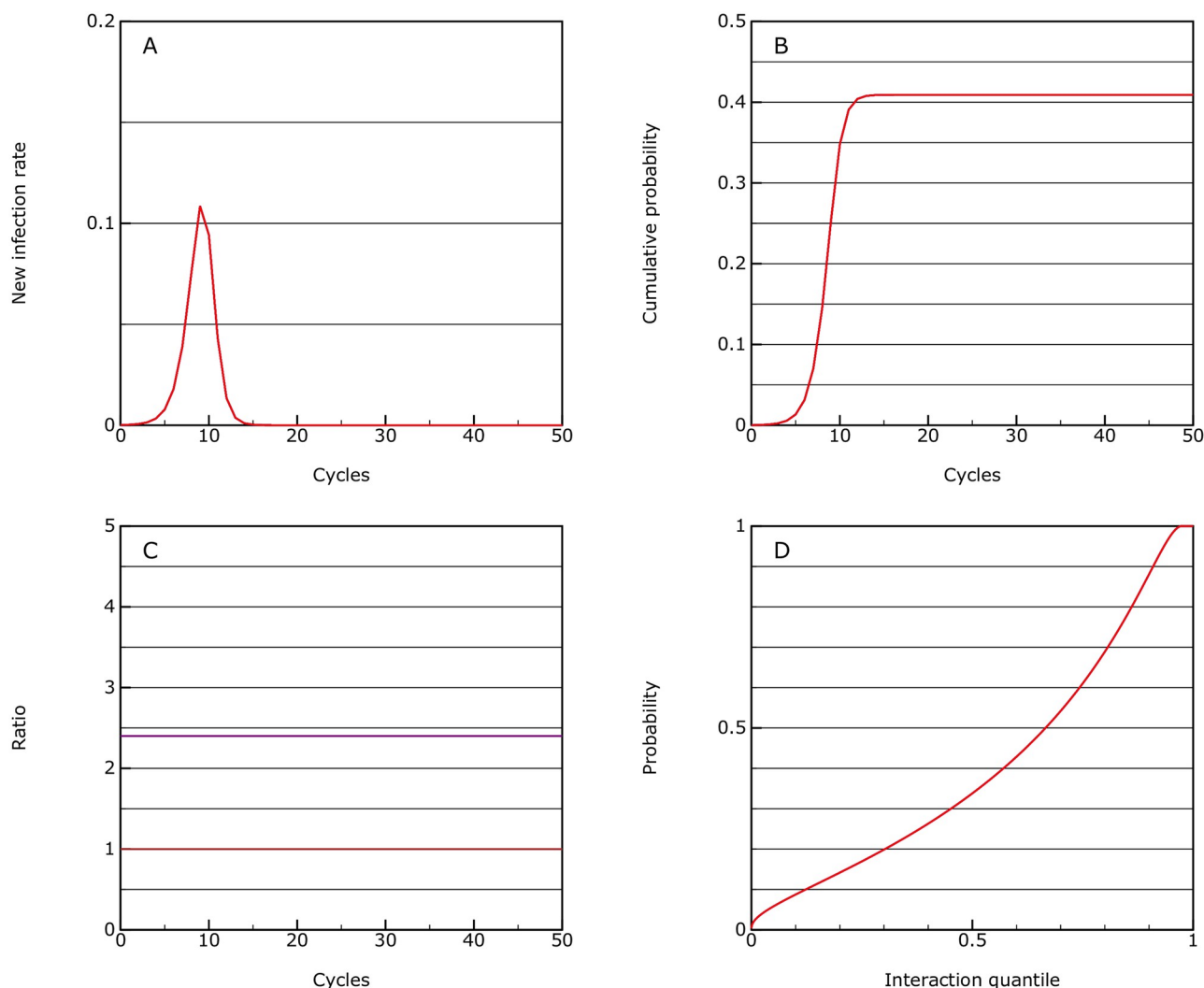
Fig 1. No intervention. This shows the model behaviour with no intervention; A: the rate of new infections as a proportion of the population in each cycle of the epidemic; B: the cumulative infection rate for the population; C: the R_{eff} value (purple) which in this case is equal to $R0$, and the allowed interaction rate compared to normal (brown), through the epidemic; D: the accumulated infection probabilities for quantiles of the population ordered by interactivity.

https://doi.org/10.1371/journal.pone.0243699.g001

The mortality over 50 cycles is 0.00037, so almost an order or magnitude lower than the protection scenario. However, the infection will keep on flaring up and if taken over 10 years, the overall mortality may even rise above the protection scenario.

The overall level of interaction between people in the population as a whole is reduced by a factor of 0.63 if activity capping is used, and by 0.43 if an activity factor is used. This pattern is replicated with many of the other scenarios, with the societal cost being far greater if overall activity factors are used rather than caps to activity levels.

Suppression. Given that it is possible to bring R_{eff} to below 1, the next scenario looks at sustaining this for longer to keep the overall case rate lower. This is achieved using sanctions at the following limits:

- 0.0002: the vulnerable group goes onto strong distancing while all others go onto limited distancing.
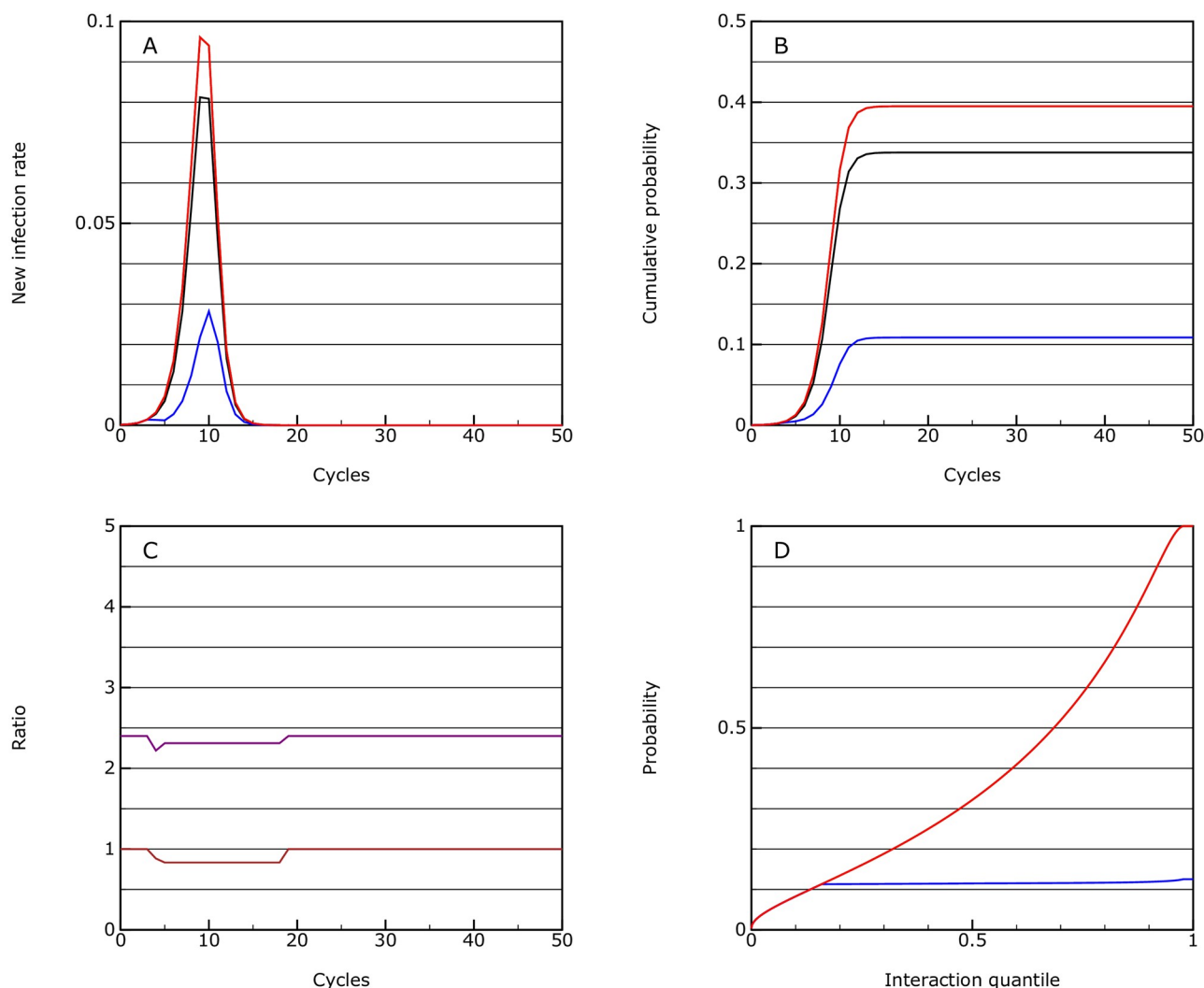
Fig 2. Protection. As for Fig 1, but where the vulnerable group (shown in blue), is protected through partial isolation from the rest of the population; the population average is shown in black.

- 0.0004: the vulnerable group goes into isolation and the main group goes into enhanced distancing.

- 0.001: the main group goes into strong distancing.

As expected the effects are similar to the mitigation strategy but the infection rate is held around the lower threshold value of 0.0004 (see Fig 4) and indeed, given the nature of the response in the model, it should be possible to stabilise at any level using this methodology (two levels of response to give a value of R_{eff} just below and just above 1).

The mortality over 50 cycles is 0.00022 with much of this being in the first pulse.

The overall level of interaction between people in the population as a whole is reduced by a factor of 0.65 if activity capping is used, and by 0.43 if an activity factor is used, essentially the same as in the mitigation scenario. There is no real penalty in stabilising the infection rate at a lower threshold.
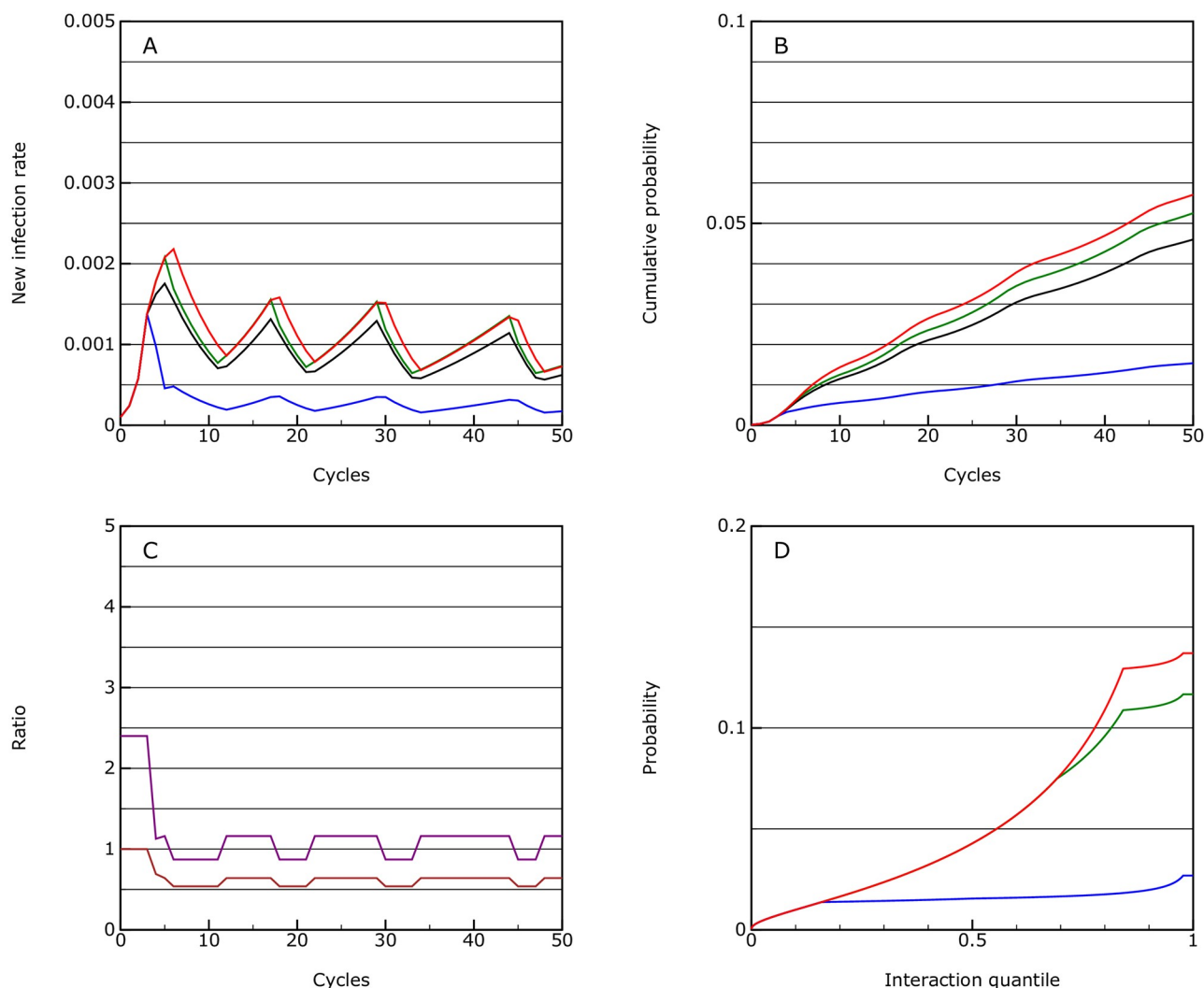
Fig 3. Mitigation. As for Fig 1, but where the vulnerable group (shown in blue), is partially isolated from the rest of the population, the majority of the population (green) has significant distancing measures applied and key workers (red) have limited distancing measures; the population average is shown in black; this is the mitigation scenario.

Lockdown. So far the scenarios have not involved putting the main group under very severe restrictions, because, according to this theoretical model, this should not be necessary to bring R_{eff} below 1 if capping of activity levels is used. However, it is worth seeing the effect of more severe restrictions is under this model and here is achieved using sanctions at the following limits:

- 0.0002: the vulnerable group goes onto strong distancing while all others go onto limited distancing.

- 0.0004: the vulnerable group goes into isolation and the main group goes into strong distancing.

- 0.001: the main group goes into isolation.
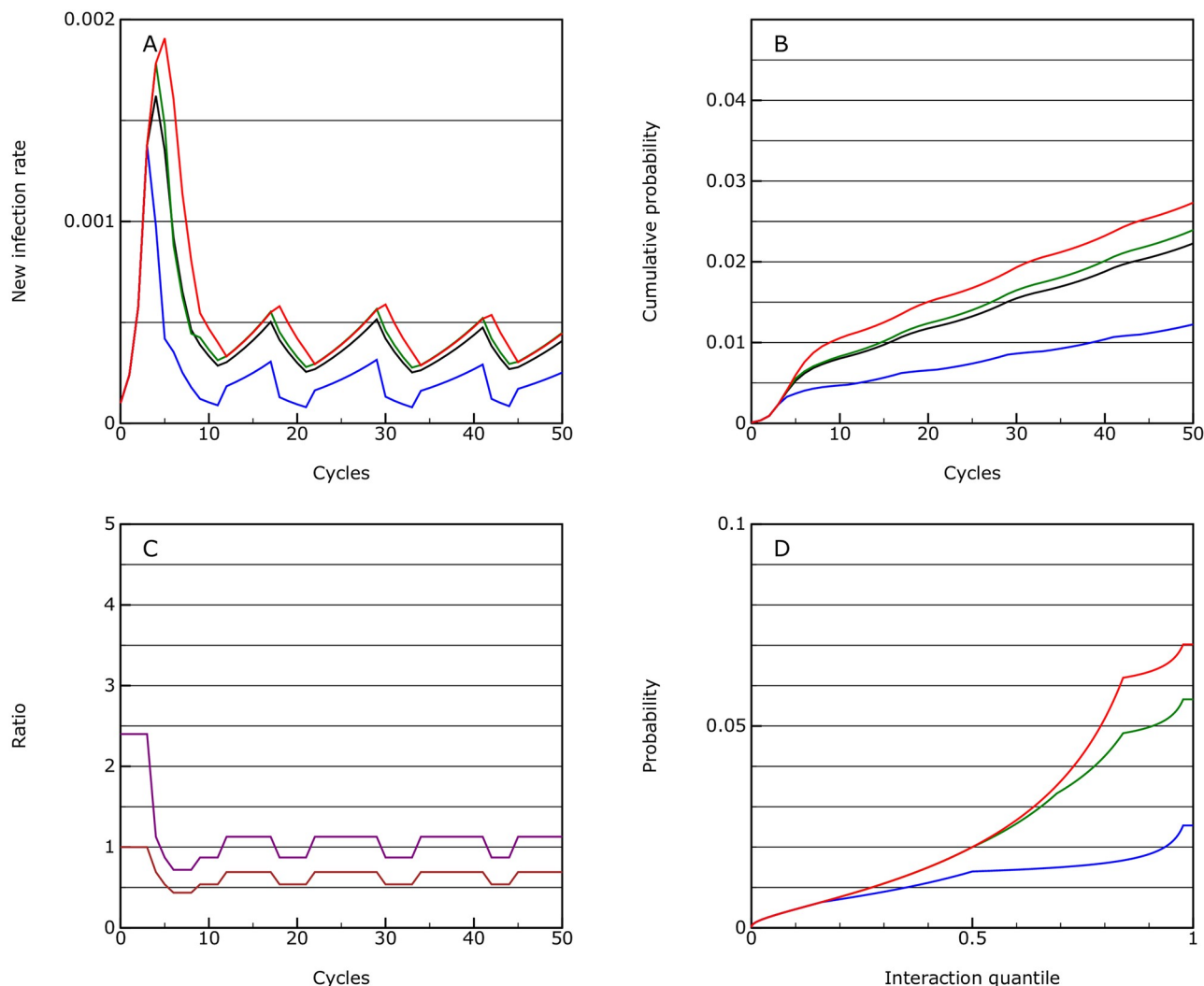
Fig 4. Suppression. As for Fig 1, but with lower thresholds for measures to be applied; this is the suppression scenario.

This leads to a less stable case level, and ends up with cycling between all three states (see Fig 5). This is because R_{eff} is less than one in the all but the lowest state. The overall activity levels remain around 0.65 (or 0.46 if a factor approach is used), but the activity levels fluctuate in a way which is likely to be very disruptive. The mortality over 50 cycles is similar at 0.00032 (if anything higher because sometimes all restrictions are lifted) and there seems to be little to be gained from this strategy from a technical point of view, but it may be necessary politically if other attempts at reducing interaction levels do not work.

Interestingly if no reduction in activity of the key worker group is made, it is very difficult to stabilise the situation, even with the majority of the population in lockdown (see Fig 6). If the key workers maintain their normal interaction levels, the epidemic will take off in that group and spread to the rest, with over 20% of this group infected by the end (and almost all of the most interactive individuals). The overall mortality rate is much higher at about 0.001 with much of that, inevitably, in the key worker group. This finding is also true under the mitigation and suppression strategies and underlines the need for some social distancing for this group
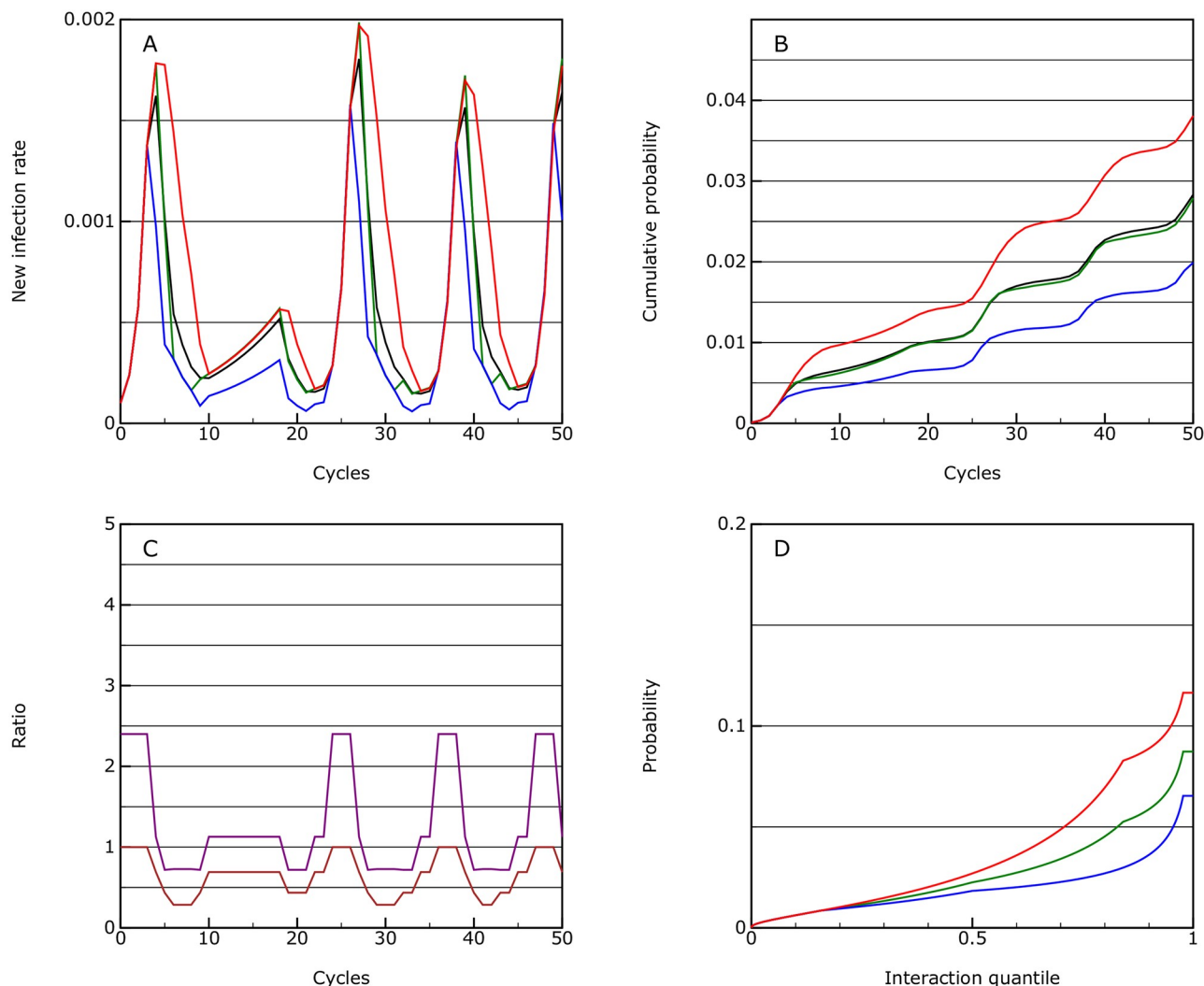
Fig 5. Lockdown. As for Fig 4, but with more extreme measures imposed on the majority of the population; this is the lockdown scenario.

(combined with personal protection and testing). The main risk in managing an epidemic, especially if initial measures are not brought in fast enough, is that the key workers become over-stretched and if anything their interaction levels rise: this positive feedback effect is something to be avoided if at all possible.

The situation presented in Fig 6 is equally valid for any other minority group which continues normal activity, either by choice or by force of circumstance. In this context it is important to note that, although the overall R_{eff} value is well above 1, the proportion of the overall population infected at the end is much lower than when there is no intervention.

Elimination. Given it is possible to reduce R_{eff} below 1, it is clearly possible to aim for elimination. This can be achieved by lowering the first threshold to zero (permanent sanctions). The measures are then set at the following levels:

- 0.0000: the vulnerable group goes onto strong distancing, the main group goes into enhanced distancing and the key workers go onto limited distancing.
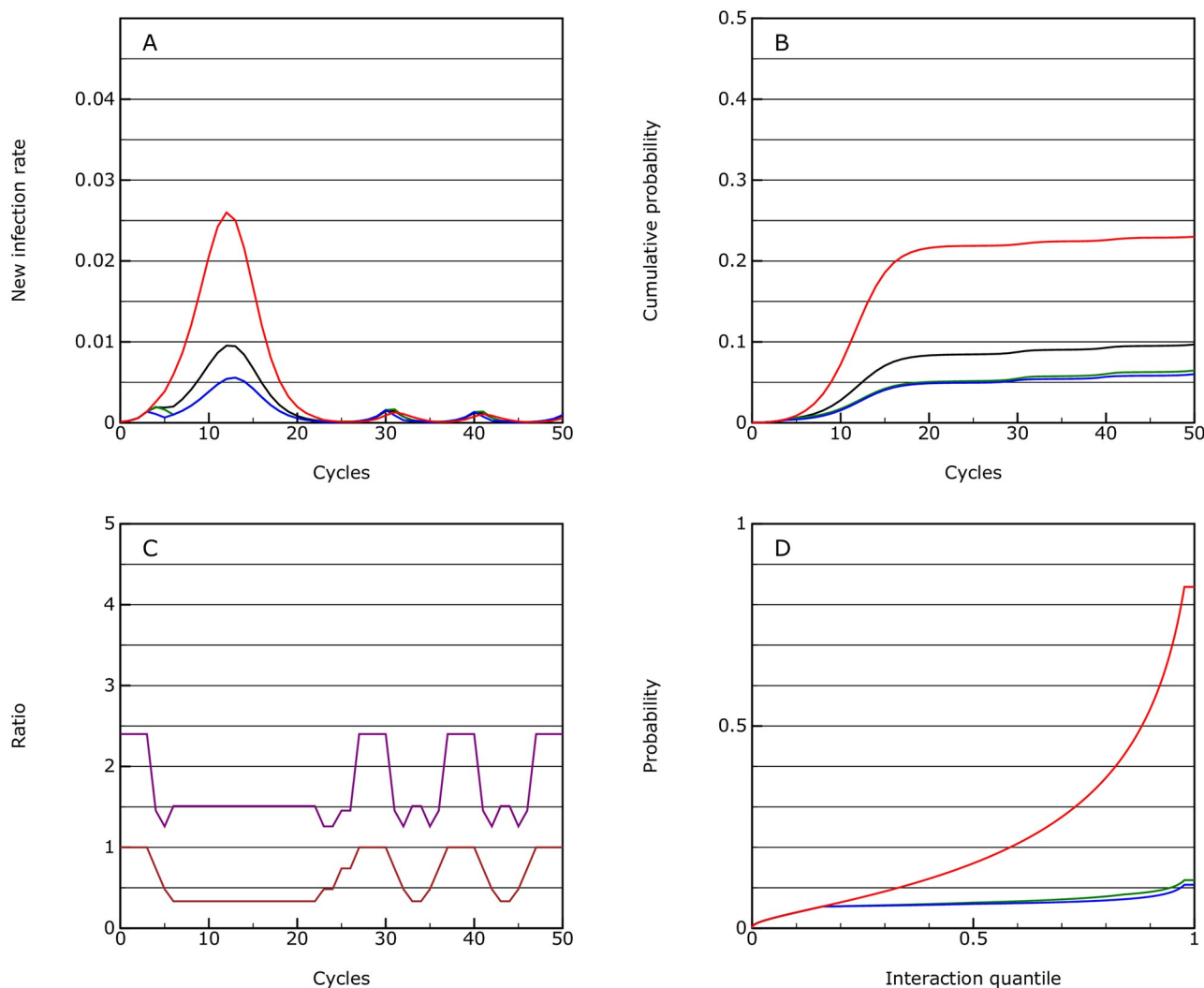
Fig 6. Lockdown with unrestricted key workers. As for Fig 5, but with the key worker group maintaining their normal interaction levels.

- 0.0004: the vulnerable group goes into isolation.

- 0.001: the main group goes into strong distancing.

This will effectively eliminate the disease (case rate to less than 1 in a million) over 50 cycles (see Fig 7). Clearly stronger sanctions would do so faster, but actually if these are not applied to the key worker group too, the differences are not that great.

The overall activity levels for the 50 cycles are 0.60 of normal levels (or 0.37 if a factor approach is used). However, the advantage here is that after those 50 cycles, potentially all restrictions could be removed.

In this scenario it takes 17 cycles to return to the seed probability. This implies that if it were possible to detect an infection rate of 1 in a million within the population and implemented the sanctions immediately, it would take about 17 weeks to eliminate the infection again. This shows that it would certainly be preferable to use good quarantining, testing and contact tracing methods [7] to prevent reinfection instead.
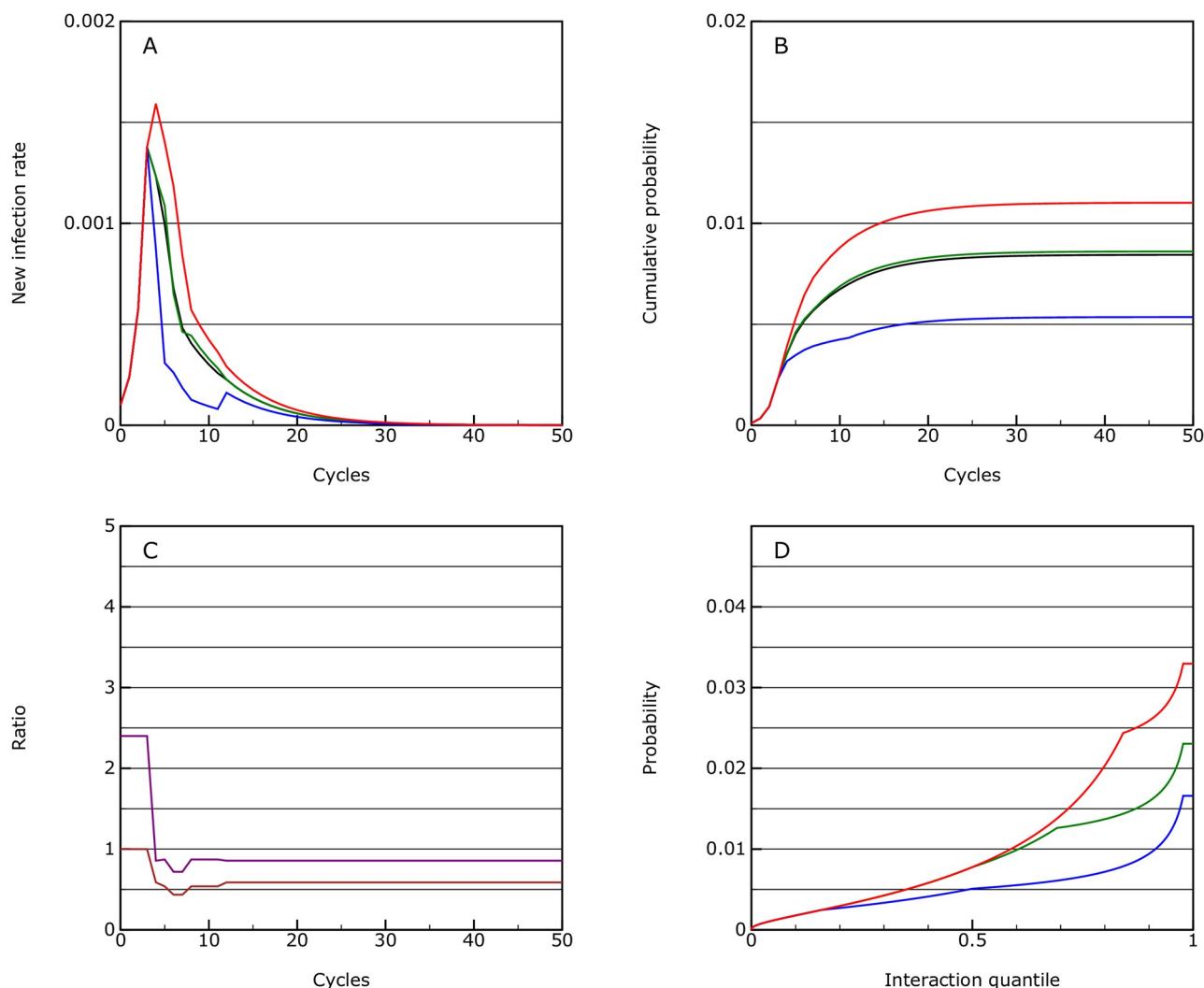
Fig 7. Elimination. As for Fig 4, but with with measures being imposed continuously until elimination.

Implications of scenario modelling. The scenario tests performed with this model do not throw up any major surprises. The protection scenario is better than the no-intervention approach, but can probably only reduce the mortality rate by a factor of about two. These options would only be justified if the societal costs of other interventions were very high.

Of the other scenarios, the societal cost in terms of lost interaction time all seem to be very similar. For everything but the elimination scenario, it is necessary to have a measured switching on and off of restrictions which keeps R_{eff} above and below 1. This will be more stable if the changes between the two states is relatively small, both in terms of controlling the epidemic and in terms of minimising disruption.

The difficulties of control make the elimination scenario particularly attractive. This allows society to adjust to a stable regime over about 50 cycles while the infection rate falls to zero. Once this has been achieved it would then be essential to prevent reinfection through robust testing and contact tracing mechanisms. It is also very likely that such mechanisms could be used at the end of the elimination process to speed up the ending of restrictions.

Conclusions

For air-borne virus infections, the strong relationship between the probability of picking up an infection and passing it on is critical to the evaluation of distancing strategies. In order to take account of this properly, it is necessary to consider the very wide range of levels of interaction that individuals have through the choices they make and the jobs that they do.

A very simple model of infection propagation demonstrates that it is much more efficient, when considering interventions to reduce the growth rate of an epidemic to below 1, to target interventions at those activities that result in the very highest interaction levels, and specifically for people who are engaged in those on a regular basis. There is a non-linearity in the system which implies that limiting the upper levels of interaction will work more effectively than reducing interactions across the population as a whole. If we can equate lost interaction time with societal cost there is something like a 50% reduction in the cost of taking this approach. Such limits in interaction might be addressed with regulation when related to work, but will require individuals to make the right choices themselves when it comes to social activity; this is where human agency becomes important in epidemic control.

The other consequence of this is that it should be possible to reduce the key $R0$ factor from a value of 2.4 to an an effective value, R_{eff}, of well below 1, without having a major impact on the majority of the population, if the interventions are properly targeted. Even used on its own, it looks as if sustainable social distancing might be able to eliminate the virus, and this will only be made more practical by other approaches such as testing, quarantining and contact tracing. On the other hand social distancing on its own is not likely to be a good mechanism for preventing re-infection; here quarantining, testing and contact tracing [7] would be much more efficient.

Another important implication of this model is that isolation of the majority of the population, while maintaining normal interaction levels for a minority of key workers, or for any other minority group, is likely to result in very high infection rates in that group and a failure to eliminate the infection within a reasonable time frame. All groups in society must engage in some social distancing, or have other protective measures, to gain effective control.

The model presented here is a purely theoretical one, based on estimated parameters. It is intended to stimulate further modelling and data collection work which can be used to test the conclusions reached here before they are used to inform policy. However, it is certainly safe to conclude that to reduce R_{eff} it is very important to reduce the interaction levels of the most interactive individuals. This is also something seen in simulation and network-based models, but the very simple probability-based model here serves to illustrate why this is case.

The outputs of the model show why it is important to include heterogeneity at the individual level in any epidemic model for COVID-19, and implies that the homogeneous assumptions that underly many models might be misleading when considering herd immunity thresholds and the effects of non-pharmaceutical interventions. The approach taken here should be simple to implement within more complex homogeneous partitioned models in widespread use, ideally with better data for estimating variability in interaction rates for different groups.

Supporting information

S1 File. The model used for the scenario testing.
(HTML)

Author Contributions

Conceptualization: Christopher Bronk Ramsey.

Investigation: Christopher Bronk Ramsey.

Methodology: Christopher Bronk Ramsey.

Project administration: Christopher Bronk Ramsey.

Software: Christopher Bronk Ramsey.

Writing – original draft: Christopher Bronk Ramsey.

Writing – review & editing: Christopher Bronk Ramsey.

References

1. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London; 2020. Available from: http://hdl.handle.net/10044/1/77482.

2. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006; 442(7101):448–452. https://doi.org/10.1038/nature04795 PMID: 16642006

3. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Rev Mod Phys. 2015; 87:925–979. https://doi.org/10.1103/RevModPhys.87.925

4. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS Medicine. 2008; 5(3):e74. https://doi.org/10.1371/journal.pmed.0050074 PMID: 18366252

5. Kissler SM, Klepac P, Tang M, Conlan AJK, Gog JR. Sparking "The BBC Four Pandemic": Leveraging citizen science and mobile phones to model the spread of disease. bioRxiv. 2018. https://doi.org/10.1101/479154

6. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of COVID-19 disease. medRxiv. 2020. https://doi.org/10.1101/2020.03.09.20033357

7. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020. https://doi.org/10.1126/science.abb6936 PMID: 32234805