medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Generic probabilistic modelling and non-homogeneity issues for the UK epidemic of COVID-19

Anatoly Zhigljavsky*, Roger Whitaker*, Ivan Fesenko†, Kobi Kremnizer‡, Jack Noonan*, Paul Harper*, Jonathan Gillard*, Thomas Woolley*, Daniel Gartner*, Jasmine Grimsley§, Edilson de Arruda¶, Val Fedorov‖, Tom Crick MBE**

April 4, 2020

Abstract

Coronavirus COVID-19 spreads through the population mostly based on social contact. To gauge the potential for widespread contagion, to cope with associated uncertainty and to inform its mitigation, more accurate and robust modelling is centrally important for policy making.

We provide a flexible modelling approach that increases the accuracy with which insights can be made. We use this to analyse different scenarios relevant to the COVID-19 situation in the UK. We present a stochastic model that captures the inherently probabilistic nature of contagion between population members. The computational nature of our model means that spatial constraints (e.g., communities and regions), the susceptibility of different age groups and other factors such as medical pre-histories can be incorporated with ease. We analyse different possible scenarios of the COVID-19 situation in the UK. Our model is robust to small changes in the parameters and is flexible in being able to deal with different scenarios.

This approach goes beyond the convention of representing the spread of an epidemic through a fixed cycle of susceptibility, infection and recovery (SIR). It is important to emphasise that standard SIR-type models, unlike our model, are not flexible enough and are also not stochastic and hence should be used with extreme caution. Our model allows both heterogeneity and inherent uncertainty to be incorporated. Due to the scarcity of verified data, we draw insights by calibrating our model using parameters from other relevant sources, including agreement on average (mean field) with parameters in SIR-based models.

We use the model to assess parameter sensitivity for a number of key variables that characterise the COVID-19 epidemic. We also test several control parameters with respect to their influence on the severity of the outbreak. Our analysis shows that due to inclusion of spatial heterogeneity in the population and the asynchronous timing of the epidemic across different areas, the severity of the epidemic might be lower than expected from other models.

We find that one of the most crucial control parameters that may significantly reduce the severity of the epidemic is the degree of separation of vulnerable people and people aged 70 years and over, but note also that isolation of other groups has an effect on the severity of the epidemic. It is important to remember that models are there to advise and not to replace reality, and that any action should be coordinated and approved by public health experts with experience in dealing with epidemics.

The computational approach makes it possible for further extensive scenario-based analysis to be undertaken. This and a comprehensive study of sensitivity of the model to different parameters defining COVID-19 and its development will be the subject of our forthcoming paper. In that paper, we shall also extend the model where we will consider different probabilistic scenarios for infected people with mild and severe cases.

*Cardiff University
†University of Nottingham
‡University of Oxford
§ONS Data Science, Newport
¶Federal University of Rio de Janeiro, Brazil
‖Icon Clinical Research, Philadelphia, USA
**Swansea University

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

1 Introduction

We model the development of the COVID-19 epidemic in the UK under different scenarios of handling the epidemic. There are many standard epidemiological models for modelling epidemics, see e.g. [1, 2]. In this research, we use a more generic simulation model which has the following features:

- it can take into account spatial heterogeneity of the population and heterogeneity of development of epidemic in different areas;
- it allows the use of time-dependent strategies for analysing the epidemics;
- it allows taking into account special characteristics of particular groups of people, especially people with specific medical pre-histories and elderly.

Standard epidemiological models such as SIR and many of its modifications do not possess these properties and hence are not applicable for any study that requires the use of the features above. In particular, In particular, for influenza the mortality changes less significantly with age in comparison to coronavirus; hence common influenza models do not give much insight in modelling the COVID-19 epidemic. The report [3], which is widely considered as the main document specifying the current epidemic in the UK and in the world, is almost entirely based on the use of standard epidemiological models and hence the conclusions of [3] seem to be lacking specifics related to important issues of the study such as heterogeneity of development of epidemic at different locations and even flexible use of different death rates across different ages. Moreover, as noted in [4], the basic assumptions for the model of [3] were written more than 13 years ago and based on the specific dynamics of a flu pandemic and hence the model was not calibrated for COVID-19. Unreliability of COVID-19 data, including the numbers of COVID cases and COVID deaths, is a serious problem. It is discussed, in particular, by G. Antes, in [5].

Despite the SIR models including the model of [3] have been heavily criticised [4, 6], when choosing the main parameters of our model we calibrate it so that its mean field version approximately reproduces the same output as SIR-based models of [3, 7] and we use the parameter values consistent with [3, 7]. Also, the main notation (which sometimes does not look natural from a statistician's point of view) is taken from [7].

The primary objective of our work is construction of a reliable, robust and interpretable model describing the epidemic under different control regimes. In this paper, we make the first step towards this objective. Due to lack of time and unreliability of available data on COVID-19, most of our results serve as an illustration of the role of different control options and we hope that even as illustration they are useful. However, we try to stay close to the COVID-19 epidemic scenario and hence we use appropriate recommendations about the choice of parameters and models of the virus behaviour we found from the studies based on the use of standard models.

See conclusions in Sections 3,4,5 for the main conclusions of this work.

Acknowledgement. We are grateful to several colleagues in Univ. Cambridge and Univ. Warwick for useful comments on drafts of this paper and to M. Hairer (ICL) and J. Ball (Univ. Nottingham) for valuable comments and discussions.

2 The model

Variables and parameters

- t - time (in days)
- t_0 - intervention time (e.g., the time when self-isolation starts)
- N - population size
- $I(t)$ - number of infected at time t
- $R(t)$ - number of immune (recovered or dead) at time t (this makes full sense from a modelling perspective, but we appreciate that this may look kind of strange when read by non-scientists)
- $S(t) = N - I(t) - R(t)$ - number of susceptible at time t
- M - number of groups selected for the study
- N_m - size of m-th group ($N = N_1 + N_2 + \ldots + N_M$)
- $I_m(t)$ - number of infected at time t in group m
- $R_m(t)$ - number of immune at time t in group m
- $S_m(t)$ - number of susceptible at time t in group m
- β - average number of transmissions of the virus per unit of time with no intervention

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

- $1/\sigma$ - average infectious period
- k - shape parameter of the Erlang distribution
- $R_0 = \beta/\sigma$ - reproductive number (average number of people who will capture the disease from one contagious person)
- R_0' - reproductive number after intervention
- p_m - probability to recover for an infected person from the m-th group

Values of parameters and generic model

The reproductive number R_0 is the main parameter defining the speed of development of an epidemic. There is no true value for R_0 as it varies in different parts of the UK (and the world). In particular, in rural areas one would expect a considerably lower value of R_0 than in London. Authors of [3] suggest $R_0 = 2.2$ and $R_0 = 2.4$ as typical; the authors of [7] use values for R_0 in the range [2.25, 2.75]. We shall use the value $R_0 = 2.5$ as typical which may be a slightly pessimistic choice overall but could be an adequate choice for the mega-cities where the epidemics develop faster and may lead to more causalities. In rural areas, in small towns, and everywhere else where social contacts are less intense, the epidemic is milder.

We assume that the person becomes infected τ days after catching the virus, where τ has Poisson distribution with mean of 1 week. To model the time to recover (or die) we use *Erlang distribution* with shape parameter $k = 3$ and rate parameter $\lambda = 1/7$ so that the mean of the distribution is $k/\lambda = 1/\sigma = 21$ (in simulations, we discretise the numbers to their nearest integers). This implies that we assume that the average longevity of the period of time while the infected person is contagious is 21 days, in line with the current knowledge, see e.g. [8, 9, 10]. Standard deviation of the chosen Erlang distribution is approximately 12, which is rather large and reflects the uncertainty we currently have about the period of time a person needs to recover (or die) from COVID-19. An increase in $1/\sigma$ would prolong the epidemic and smaller values of $1/\sigma$ would make it to cause people to be contagious for less time. The use of Erlang distribution is standard for modelling similar events in reliability and queuing theories, which have much in common with epidemiology. We have considered the sensitivity of the model in this study with respect to the choice of parameters λ and σ but more has to be done in cooperation with epidemiologists. As there are currently many outbreaks epidemics, new knowledge about the distribution of the period of infection by COVID-19 can emerge soon.

The model varies depending on purposes of the study. The main ingredients are: $I_m(t)$ is the birth-and-death processes and $R_m(t)$ is the associated pure birth processes. The process of transmission is the Poisson process with intensity β (time to next transmission has the exponential density $\beta e^{-\beta t}$, $t > 0$). After the intervention (for $t \geq t_0$), the Poisson process of transmission for m-the group has intensity β_m.

We treat this model as purely stochastic despite parts of it can be written it terms of systems of stochastic differential equations. Despite running pure simulation models taking longer than running combined models, they are simpler and less prone to certain errors.

The number of infected at time t as the main quantity of interest

To start with, in Figure 1 we consider an uninterrupted run of an epidemic with $R_0 = 2.5$ in a homogeneous (one-group) population. The starting time of an epidemic is unknown and is even hard to define as the first transmissions of the virus take random and perhaps long times. We start plotting the curves after 0.5% of the population are infected.

In red colour, in Figure 1 and all plots below we plot values of $S(t)/N$, the proportion of people non-infected by time t. In blue, we plot $R(t)/N$, the proportion of people recovered from the disease (or dead) and in green we plot our main quantity of interest which is $I(t)/N$, the proportion of infected people at time t. The values of $R(t)$ do not play any part in modelling and are plotted for information only.

The duration of time while a person with a virus is infected is modelled by Erlang distribution with shape parameter $k = 3$ and mean 21. The simulations are flexible and we can easily change these values and plot similar curves with updated data. Despite the full recovery taking slightly longer, $1/\sigma = 21$ days is a good estimation for the period of time when people with severe infection may require intensive care, ventilator etc. Also, we believe that this is a suitable distribution for the period of time when such people may die.

From the value $I(t)$, we can estimate the distribution of the number of deaths, at time t, as follows: first, since on average an infected person is considered as infected for 21 days, we compute $I(t)/21$. At each particular day, any person can die with the probability which is typical for the chosen population or sub-population. If we consider the whole population, then we apply the mortality rate for the population. However, if $I(t)$ refers to a particular group only, then we should apply the corresponding coefficient for

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

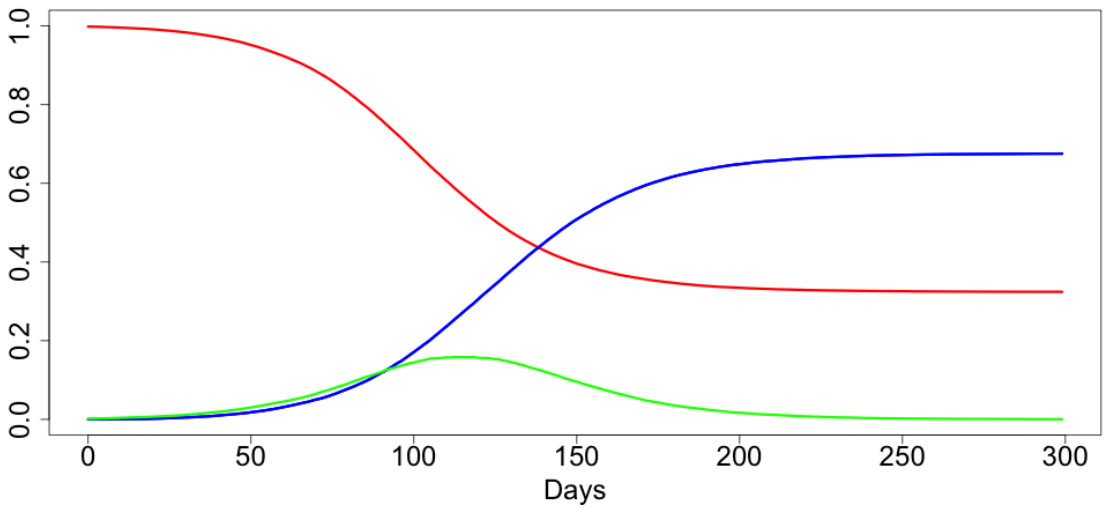

Figure 1: An uninterrupted run of a COVID-19 epidemic in homogeneous conditions

the group. In any case, the distribution for the number of deaths at time t can be roughly considered as binomial with parameters $I(t)$ and the probability of 'success' $r\sigma$, where r is the mortality rate. Different authors disagree on the values of mortality r for the COVID-19; see, for example, [11]. UK's experts believe $r \simeq 0.009$ [3], WHO sets the world-wide mortality rate at 0.034, the authors of [12] believe r is very small and could be close to 0.001, an Israeli expert D. Yamin sets $r = 0.003$, see [13].

Example 1. *Assume that the epidemic in Birmingham (with population size $N \simeq 1,086m$) was running uninterrupted according to the scenario depicted in Figure 1. The maximum value of $I(t)/N$ is $\simeq 0.163$ giving the maximum expected daily death toll of $0.163 \cdot 1086000 \cdot 0.009/21 \simeq 75.8$ assuming we use UK experts value $r = 0.009$.*

Although simple to implement, the estimation in Example 1 is likely to be wrong, as it would be for any heterogeneous city (e.g. London). Critically, as will be discussed in the next section, the maximum death toll is likely to be significantly lower than the above calculations suggest.

Summarizing, the expected daily mortality curve is simply a suitably scaled version of $I(t)/N$ of Figure 1. The same is applied to the curves representing expected number of hospital beds and ventilators. The variability for the number deaths is naturally considerably higher than the variability for the numbers of required hospital beds and ventilators as the latter ones are correlated in view of the fact that each person occupies a bed (requires a ventilator) for a few days in a row but dies only once.

Unlike $I(t)$, the number of deaths $D(t)$ at each particular day is (approximately) known and potentially can be used for estimating the first and second derivatives of $I(t)/N$ and hence the stage of the epidemic. It is, however, difficult to do for the following reasons: randomness in the values $D(t)$, see above, and, more importantly, heterogeneity of large sub-populations, see next section.

One of the main targets of decision-makers for dealing with epidemics like COVID-19 can be referred to as 'flattening the curve', where 'the curve' is $I(t)/N$ or any of its equivalents and 'flattening' roughly means 'suppressing the maximum'. We shall consider this in the next sections.

3 Spatial heterogeneity of the population and heterogeneity of epidemic development in different areas

The purpose of this section is to demonstrate that heterogeneity of epidemic developments for different sub-populations has significant effect on 'flattening the curve'.

Already when this paper was completed, the authors learned about a preprint [14] which uses SIR modelling to produce somewhat similar conclusions with more specifics for COVID-19 in the UK. However, SIR modelling using numerous age-classes and many equations for each age-class different across different parts of the country resulting in an astronomical number of parameters; this makes any sensitivity analysis to various parameters uncheckable. We believe that a combination of stochastic models like the present one with standard SIR-based models can help in significantly reducing the number of parameters while also adding more flexibility to a hybrid model.

Consider the following situation. Assume that we have a population consisting of M sub-populations (groups) G_m with similar demographic and social characteristics and these sub-populations are subject

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

to the same epidemic which has started at slightly different times. Let the sizes of sub-population G_m be N_m with $N_1 + \ldots + N_m = N$. In all cases, the curves $I_m(t)/N_m$ are shifted in time versions of the curve $I(t)/N$ of Figure 1.

In Figures 2 and 3, $M = 2$ and the second epidemic started 50 days after the first one (incubation period is set to be 0). In Figures 4 and 5, $M = 10$ and each next epidemic cycle has started 7 days after the previous one. In all cases, we evidence the significant 'flattening the curve' phenomenon. In Figures 2–5, the green line is used for the resulting curves $I(t)/N$. The maximal values of $I(t)/N$ in the examples depicted in Figures 2–5 are 0.124, 0.136, 0.134 and 0.132, respectively. This is significantly lower than the value 0.163 for the original curve. In the assumptions of Example 1, these would respectively lead to the maximum expected number of deaths 57.7, 63.3, 62.9 and 62.7.

Extra heterogeneity of different epidemics caused by the social and demographic heterogeneity of the sub-populations would further 'flatten the curve'. The following conclusions are in line with what the Israeli expert D. Yamin states [13] and agree with the main conclusions of the paper [15] which is rather critical towards standard epidemiological models.

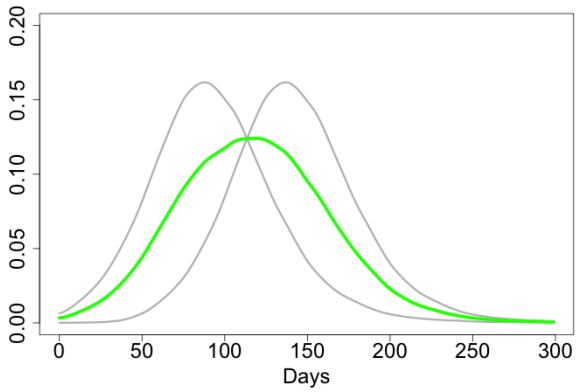

Figure 2: $M = 2$, $N_1 = N_2$.

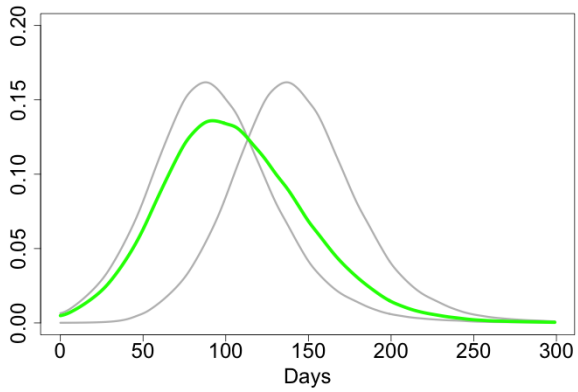

Figure 3: $M = 2$, $N_1 = 3N_2$

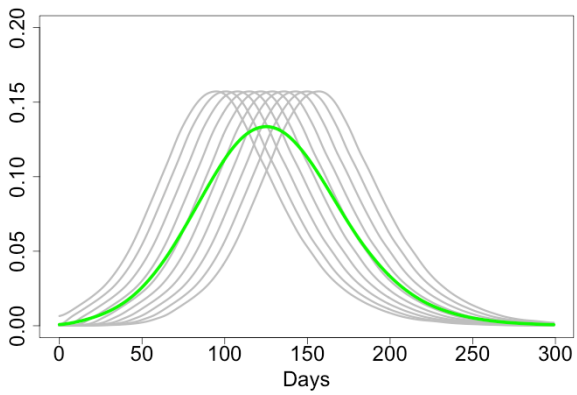

Figure 4: $M = 10$, $N_m = \frac{N}{10}$ $(m = 1, \ldots, 10)$

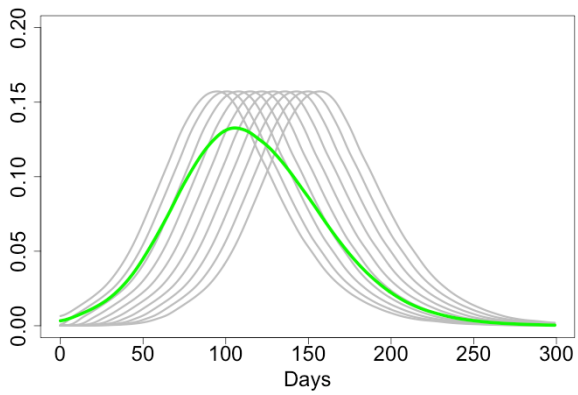

Figure 5: $M = 10$, $N_1 = \frac{N}{2}$, $N_m = \frac{N}{18}$ $(m > 1)$

Conclusions. *(a) Isolation of sub-populations at initial stages of an epidemic is very important for preserving heterogeneity of the epidemic and 'flattenning the curve'; this flattening can be very significant. (b) Epidemiological models based on the assumption of a homogeneous population but applied for populations consisting of heterogeneous sub-populations may give completely misleading results.*

4 An epidemic with intervention

In this section, we assume that we make an intervention to the epidemic by introducing an isolation at a certain stage. Moreover, we assume that isolation may be different for 2 different groups. We define a special group G of more vulnerable people consisting, for example, of all people aged 70+. We define $\alpha = n/N$, where N is the total population size and n is the size of this special group G.

Let t_0 be the moment of time when the isolation occurs. It is natural to define t_0 from the condition $S(t_0)/N = x$, where, for example, $x = 0.9$. For $t < t_0$, the virus has been transmitted to people uniformly

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

so that, conditionally a virus is transmitted, the probability that it reaches a person from group G is α. For $t \geq t_0$, the virus has been transmitted to people in such a way that, conditionally a virus is transmitted, the probability that it reaches a person from G is $p = c\alpha$ with $0 < c \leq 1$. Moreover, at time t_0 the value of R_0 may change to R_0' in view of self-isolation. The parameters in this model are:

- $\alpha = n/N$: relative size of the group G;
- $x \in (0, 1)$ defines the start of the isolation strategy;
- $c = p/\alpha$ defines the strength of isolation of the group G;
- R_0 (initial);
- R_0': the reproductive number after intervention for $t \geq t_0$; it defines the strength of the overall isolation.

We have run a large number of scenarios coded using a combination of R and Julia [16]; the code is provided in Appendix. In Figures 6–11, we illustrate a few of these scenarios. In all these scenarios, we have chosen $R_0 = 2.5$ and $\alpha = 0.2$. The results are robust towards values of these parameters (subject to faster or slower rate of the epidemic in dependence on R_0). In Section 5 we use $\alpha = 0.132$ to illustrate some specific results; $\alpha = 0.2$ can be considered either as a generic value or, in the contents of Section 5, as the relative size of the group of vulnerable people which is larger than the group of people aged 70+.

We would like to emphasise that there is still a lot of uncertainty on who is vulnerable. It is very possible that the long term effect of the virus might cause many extra morbidities in the coming years coming from severe cases who recover. It is also possible that the virus will mutate and the new strain might affect younger populations more than the current strain. These, and many more possibilities, have a non-negligible probability of occurring, and they have devastating effects. In future models we will try to address these possibilities.

To estimate the number of COVID cases at a given time t is difficult [5]. This implies that at the time of making an intervention only very rough guesses about the value of x, which is crucial for the future development of the epidemic, can be made.

We distinguish different strengths of isolation by values of parameters c and R_0':

$$c = \begin{cases} 1 & \text{no extra separation for people in } G \\ 0.5 & \text{mild extra separation for people in } G \\ 0.25 & \text{strong extra separation for people in } G \end{cases}$$

$$R_0' = \begin{cases} R_0 & \text{no social distancing requirement for general public} \\ 1.5 & \text{social distancing but no self-isolation requirement for general public} \\ 1 & \text{self-isolation requirement for general public} \end{cases}$$

The values $c = 0.5$ and $c = 0.25$ mean that under the condition that a virus is infecting a new person, the probabilities that this new person belongs to G are $1/3$ and $1/5$ respectively.

Measuring the level of compliance in the population and converting this to simple epidemiological measures c and R_0' is hugely complex problem which is beyond the scope of this paper.

The lines and respective colours in Figures 6–11 are as follows.

- Solid blue line: $R(t)/N$, the proportion of people recovered from the disease (or dead) in case of no intervention, as in Figure 1.
- Dashed blue line: $R(t)/N$ in case of intervention.
- Solid black line: $R_G(t)/n$ the proportion of people from group G recovered from the disease (or dead) in case of intervention.
- Solid red line: $S(t)/N$, the proportion of people non-infected by time t in case of no intervention, as in Figure 1.
- Solid orange line: $S_G(t)/n$, the proportion of people susceptible to the virus from group G non-infected by time t in case of intervention.
- Solid green line: $I(t)/N$, the proportion of infected people at time t with no intervention, as in Figure 1.
- Dashed green line: $I(t)/N$, the proportion of infected people at time t with intervention.
- Solid purple line: $I_G(t)/n$, the proportion of infected people from group G at time t with intervention.

Considerably important, in view of the discussions of the next section, is Figure 6. In this figure, for the initial data of Figure 1, we strongly separate group G at the time when 10% of the population is infected. We make no call to the general public for social distancing. In this scenario, the intervention does not considerably change the proportion of infected in the total population but it significantly 'flattens the curve' for the group G: compare the purple and green lines. The maximum of $I_G(t)/n$ is 0.086 which is 2.5 times lower than 0.216, the maximum of $I(t)/N$. Note also the fact that the curve $I_G(t)/n$ is

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

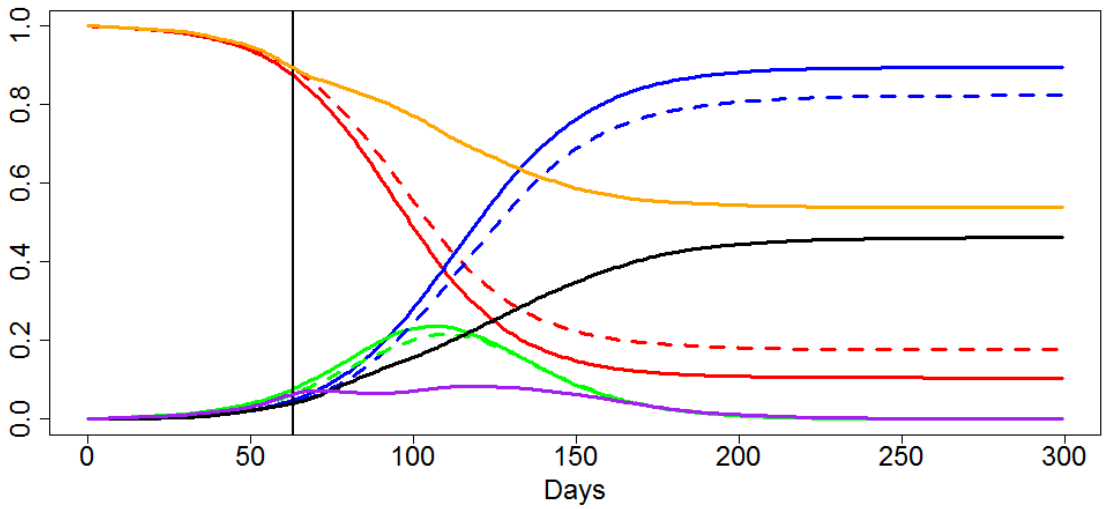

Figure 6: $x = 0.9$, $c = 0.25$, $R_0' = 2.5$

rather flat for long time, during the main stage of development of the epidemic. Just after the peak of the epidemic in the whole population, $I_G(t)/n$ peaks; it is caused by the presence of very large number of infected people from the general population.

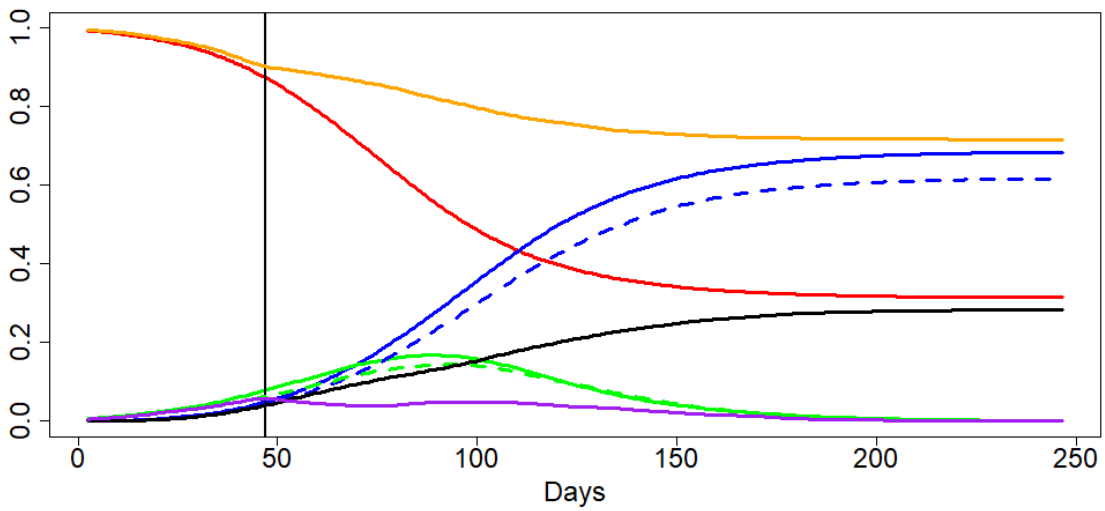

Figure 7: $x = 0.9$, $c = 0.25$, $R_0' = 2.5$, no incubation period

The scenario which led to Figure 6 is important and hence it has been re-run for a few variations of the model. Figure 7 shows the results of the simulations where we have removed the incubation period. The maximum of $I_G(t)/n$ is now 0.075 which is more than twice lower than 1.63, the maximum of $I(t)/N$. Note that in the scenario with no incubation period the whole epidemic is milder as the virus lives longer.

In Figure 8 we use similar scenario as for Figure 6 but we separate group G only mildly at the time when 10% of the population is infected. We make no call to general public for social distancing. The curve $I_G(t)/n$ is 'flattened' for the group G but in a considerably smaller degree than in the case of strong separation of G.

Figures 9 and 10 illustrate the situation with mild and strong separation of people from G complemented with introduction of the social distancing for general public. Interestingly enough, the effect of social distancing for general public gives less benefit than even mild isolation of people from group G.

Figure 11 illustrates the scenario with no call to the general public for social distancing but with strong separation of the people from G at at a later stage of epidemic, when 20% of the population is infected. The effect is similar to the one observed in Figures 6 and 7 except for the fact that the call for isolating the group G came slightly late.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

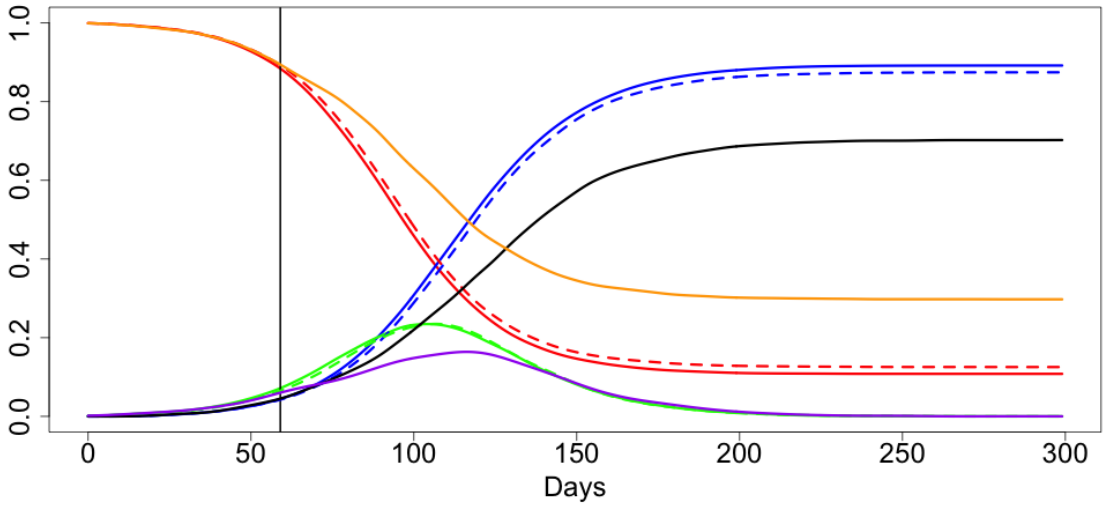

Figure 8: $x = 0.9$, $c = 0.5$, $R_0' = 2.5$

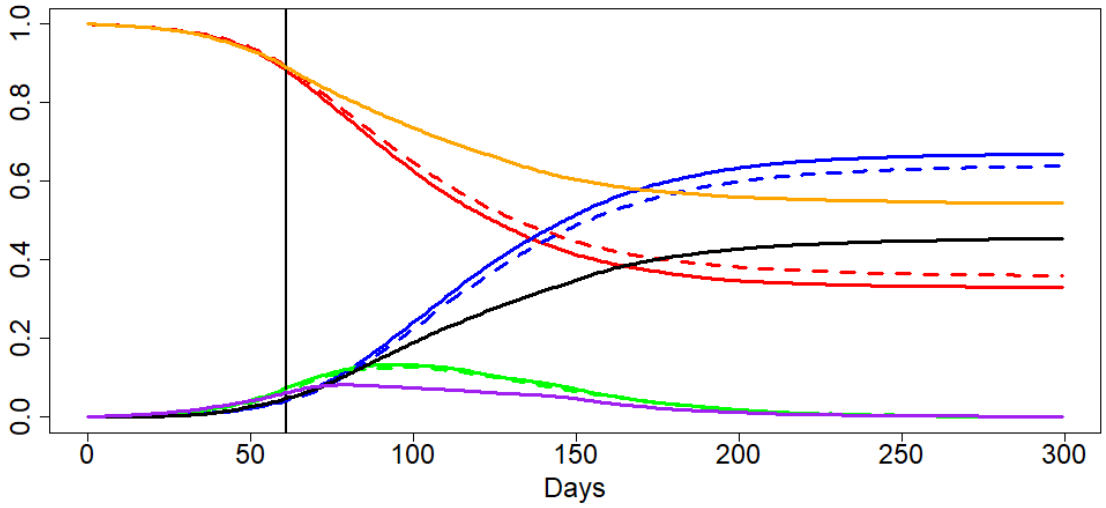

Figure 9: $x = 0.9$, $c = 0.5$, $R_0' = 1.5$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

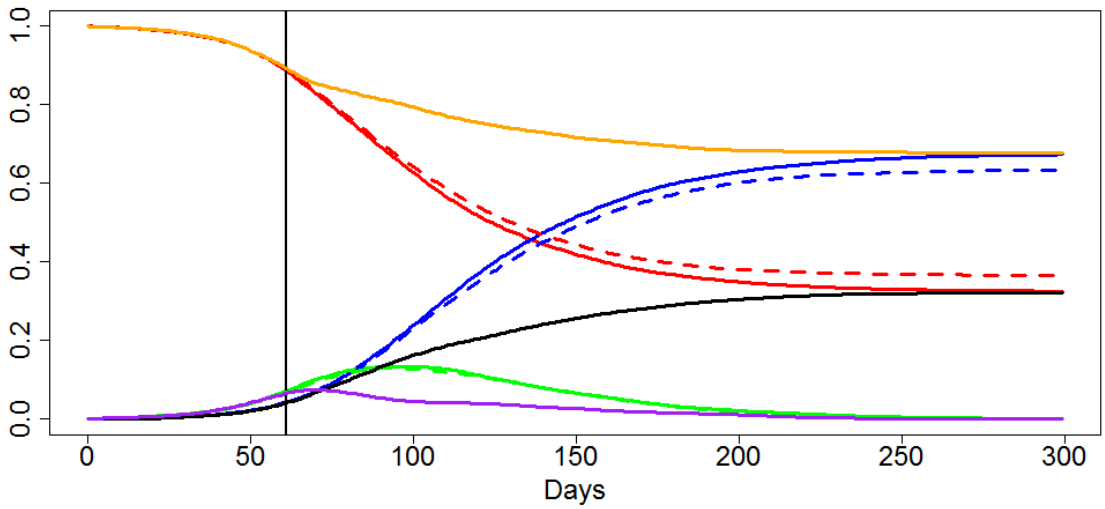

Figure 10: $x = 0.9$, $c = 0.25$, $R_0' = 1.5$

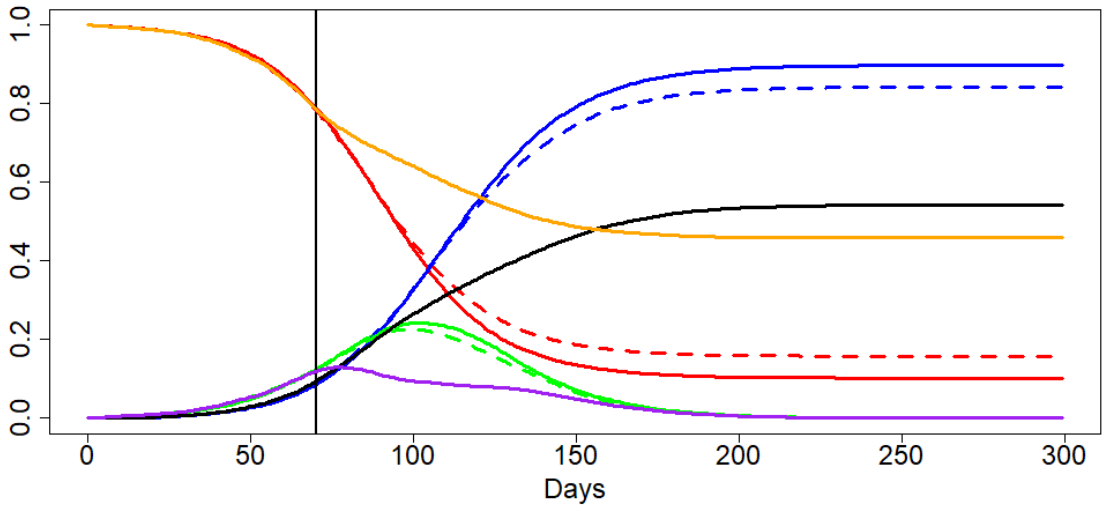

Figure 11: $x = 0.8$, $c = 0.25$, $R_0' = 2.5$

Conclusion. *By considering a number of scenarios we have observed an extreme sensitivity of the negative consequences of the epidemic to the degree of separation of vulnerable people. Sensitivity to the parameter measuring the degree of self-isolation for the whole population is less apparent although it is much costlier.*

This conclusion is very much in line with recommendations of leading German epidemiologists [17].

5 Consequences for the mortality and impact on NHS

Results of the type presented in Figures 6–11 can be translated into the language of the expected number of death and expected number of beds required. To do this we extend the observations we made at the end of Section 2 concerning translation of the curve $I(t)/N$ into the curves for the expected number of death $ED(t)$ and expected number of hospital beds in the UK. We use the common split of the UK population into following age groups:

$$G_1 = [0, 19],\ \ G_2 = [20, 29],\ G_3 = [30, 39],\ G_4 = [40, 49],$$
$$G_5 = [50, 59],\ G_6 = [60, 69],\ G_7 = [70, 79],\ G_8 = [80, \infty)$$

and corresponding numbers N_m ($m = 1, \ldots, 8$; in millions) taken from [18]

$$[15.58, 8.71, 8.83, 8.50, 8.96, 7.07, 5.49, 3.27] \ \text{ with } N = 66.41$$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

The death probabilities p_m are given from Table 1 in [3] and replicated many times by the BBC and other news agencies are:

$$[0.00003, 0.0003, 0.0008, 0.0015, 0.006, 0.022, 0.051, 0.093]\,.$$

Unfortunately, these numbers do not match the other key number given in [3]: the UK average mortality rate which is estimated to be about 0.9%. As we feel the value of the UK average mortality rate is more important, we have multiplied all probabilities above by 0.732 to get the average mortality rate to be 0.9%.

Defining the group G as a union of groups G_7 and G_8, we have $n = 5.49 + 3.27 = 8.76$m and $\alpha = 8.76/66.41 \simeq 0.132$. We then compute the mortality rate in the group G and for the rest of population by

$$r_G = \frac{N_7 p_7 + N_8 p_8}{n} \simeq 0.049; \quad r_{other} = \frac{N_1 p_1 + \ldots + N_6 p_6}{N - n} \simeq 0.0030.$$

For estimating the average number of death in the group G and for the rest of population we can the use the formulas

$$ED_G(t) = r_G \sigma I_G(t) \;\text{ and }\; ED_{other}(t) = r_{other}\sigma(I(t) - I_G(t))$$

respectively. The average numbers of hospital beds required for two different groups are proportional to these numbers.

We have run a series of scenarios for an UK epidemic without taking into account spatial and social heterogeneity of the society assuming that we would have separated (mildly and strongly) the group G of 70+ people at the time t_0 when 10% of the population were infected with no call to general public for social distancing. We marked t_0 as March 23, 2020.

The two scenarios (with $c = 0.5$ and $c = 0.25$) we have used for illustrating this technique are different from the scenarios used for plotting Figures 6 and 7 only by the value of α. For Figures 6 and 7, we have chosen $\alpha = 0.2$ but for the group G of 70+ old people in the UK we have $\alpha = 0.13$.

In Figures 12 and 13 we plot (using the same colours as in Figures 6 and 7) the following curves: $I(t)/N$ (solid green, no intervention on March 23), $I(t)/N$ (dashed green, intervention on March 23) and $I_G(t)/n$ (purple, intervention on March 23).

In Figures 14 and 15 we plot the curves for the estimated average number of deaths: $ED(t)$ (solid green, no intervention on March 23), $ED_G(t)$ (purple, intervention on March 23), $ED_{other}(t)$ (dashed green, intervention on March 23) and combined $E[D_G(t) + D_{other}(t)]$ (black, intervention on March 23).

We can deduce from Figure 15 (taking into account the extra factor of hospital beds availability) that strong separation of 70+ old people alone would have reduced the expected number of death by at least the factor of 2. Another feature of the scenario with $c = 0.25$ is a roughly 50/50 split between the number of deaths in the 70+ and 70− groups.

The y-axis in Figures 14 and 15, after multiplication by 140, can be roughly interpreted as hundreds in the London epidemic assuming $R_0 = 2.5$, $x = 0.9$ on March 23 and homogeneity of the epidemic (as there is about 9m people in London out of 64.1m). As mentioned in Section 3, deaths numbers in other regions of the UK should be expected to be lower.

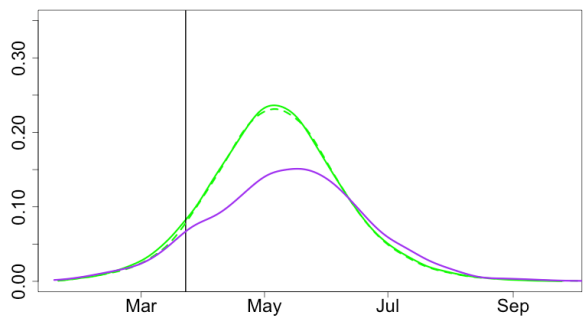

Figure 12: $c = 0.5$

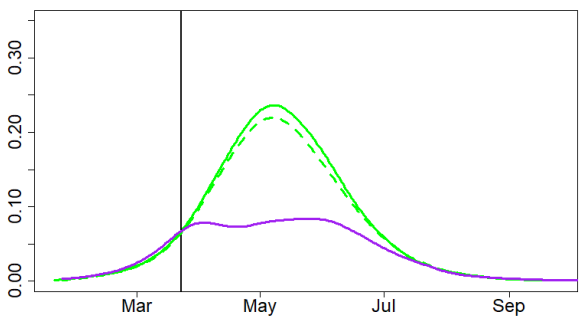

Figure 13: $c = 0.25$

We then have run through the scenarios with partial lockdown on March 23 reducing $R_0 = 2.5$ to $R_0' = 1$ and with a return, 30 days later, to $R_0 = 2.5$ with $c = 0.5$ and $c = 0.25$. Results are plotted in Figures 16-19. The style of Figure is exactly the same as for Figures 12-15. If the value x (proportion of non-infected people on March 23) happens to be larger than 0.9 then the second wave of epidemic should be expected to be (perhaps, considerably) larger. If $x < 0.9$ then the second wave will be less pronounced.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

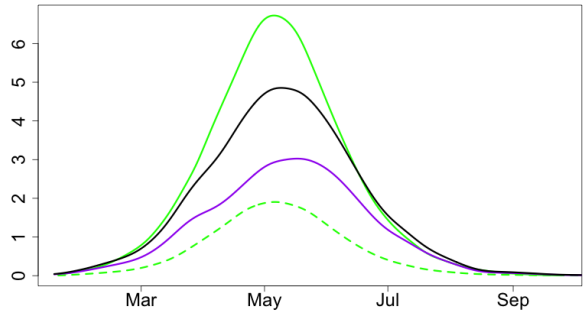

Figure 14: $c = 0.5$

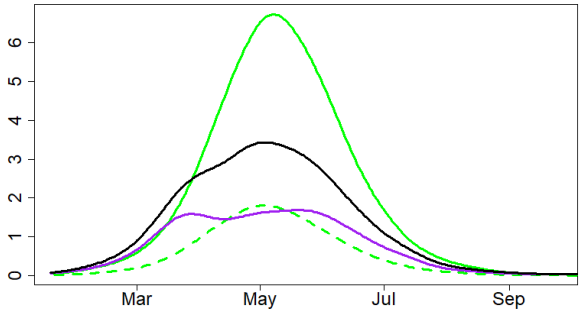

Figure 15: $c = 0.25$

All these figures are given for illustration only without any claim on accurate predictions as there is an uncertainty of the outputs towards the choice of several parameters describing the virus characteristics but the sensitivity of the model towards the choice of these parameters is not yet adequately assessed. Figures 20-21 illustrate sensitivity of the scenario of Figures 16-17 with respect to the choice of k, the shape parameter of the Erlang distribution defined in Section 2. In these figures, we plot 100 trajectories similar to the trajectories of Figures 16-17 with the values of k taken at random in the interval $[2, 4]$.

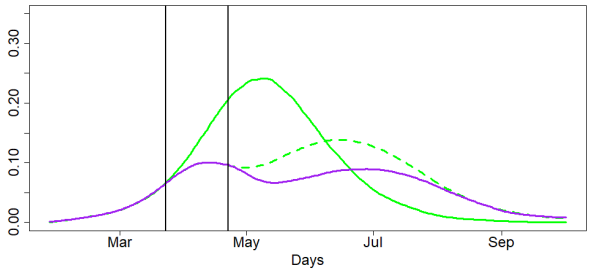

Figure 16: $c = 0.5$

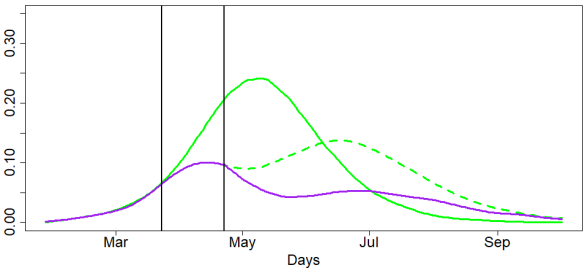

Figure 17: $c = 0.25$

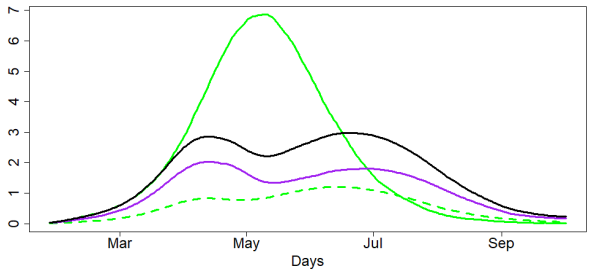

Figure 18: $c = 0.5$

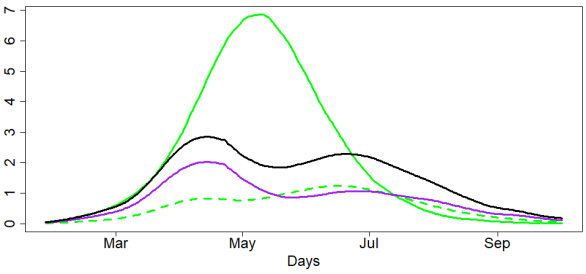

Figure 19: $c = 0.25$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

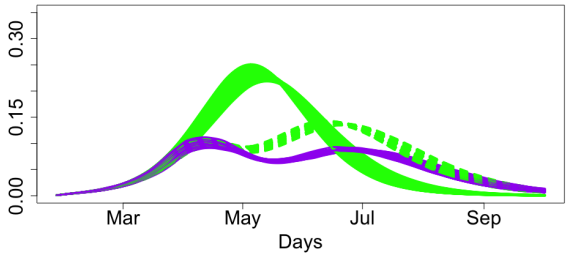

Figure 20: $c = 0.5$, $k \in [2, 4]$

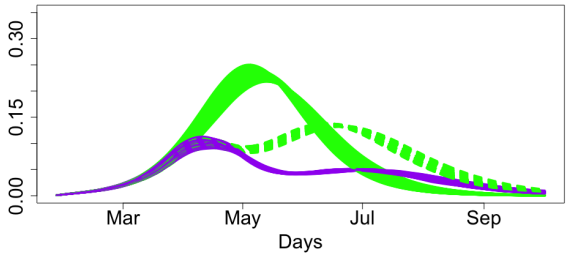

Figure 21: $c = 0.25$, $k \in [2, 4]$

Conclusion. *The outputs of the model developed above can be translated into the language of the expected number of death due to the epidemic. This language amplifies the findings of the previous section stating that isolation of a relatively small percentage of population may hugely reduce the death toll of the epidemic.*

Some of the next steps and open problems

- Run many different scenarios to get better understanding of the current situation with the epidemic and what can be done to effectively control it.
- Continue calibrating the model against existing epidemiological models and new data as it emerges.
- By using methods of stochastic global optimization [19, 20], teach the model to learn from the data emerging daily, for adapting values of parameters describing the virus and hence the course of the epidemic.
- Incorporate into the model algorithms for quickest on-line detection of changes [21, 22] to learn about temporal and spatial heterogeneity of the development of the epidemic.
- Understanding the reproductive number R_0 in dependence on the population density in local areas and hence the mixing distribution for R_0 needed to combine sub-populations.
- Understanding R_0 and other virus parameters as random variables due to mutations.
- Understanding the risk as a function depending on different factors such as age and social groups.
- Use of OR models for studying NHS capacity issues under different scenarios and correlated to the percentage of health providers getting infected by the virus.
- Properly quantify uncertainty for model predictions.
- Full-scale sensitivity analysis to different parameters.
- With better understanding of the role of parameters, formulate inverse problems like finding the stage of an epidemic by its early development.
- Use game theory [23] to produce tools for optimal decision making with respect to slowing down the process of epidemic by enforcing spatial isolation and isolation of different sub-populations.
- Develop a more sophisticated model combining stochastic differential calculus [24], fractional diffusion [25] and elements of direct simulation (the models of fractional diffusion will be used to understand the effects of slowing down of the epidemic).
- Understand and quantify the differences in epidemic developments across different countries such as UK, USA, Italy and Spain.
- Studying long-term effects of coronovirus and future mortality from Covid-19.
- Studying mortality from other diseases due to the capacity/fragility of the health system during the epidemic.
- Modelling epidemic control strategies based on testing, tracing and isolation.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Appendix: Julia/R code used for computing the scenarios of Section 4

```
#### Main script with R and Julia ####

## Running code will require installation of Julia
##install.packages("JuliaCall")
library('JuliaCall')
## In the below, specify where Julia is located
julia <- julia_setup(JULIA_HOME= "/Applications/Julia-1.4.app/Contents/Resources/julia/bin")
## Install required packages in Julia and import into R
julia_library("Random")
julia_library("Distributions")

## Main simulation

## Set total time to run simulation
Total_time=250

## Below is the main procedure. Refer to paper for description of what parameters are.

julia_command("function Main_procedure(Time,R,R_prime,N,c,quantile,alpha,sigma)
              Time=Int(Time)
              N=Int(N)

              ## Set parameters for study
              beta = R*sigma              ## Transmission coefficient
              k=3                         ## parameter for Erlang Distribution
              alpha_prime = c*alpha       ## New alpha
              beta1 =R_prime*sigma        ## New transmission coefficient if R0 changes
              Infectious_parameter=1/7    ## Average of incubation period

              ## indexes of columns in the matrix 'Population'
              person=1
              age=2
              Group=3
              Susceptible=4
              Infected=5
              Recovered=6
              Time_til_infected=7
              time_of_next_inf=8
              time_of_recovery=9
              previous_infection=10

              ## Create initial population

              G1_split = Int(N*(1-alpha))   ## Number of people in first group
              G2_split = N-G1_split         ## Number of people in second group

              Population =zeros(Int32,N,10)
              Population[:,person]=[1:N;]
              Population[:,Susceptible].=1
              Population[1:G1_split,Group].=1
              Population[G1_split+1:N,Group].=2

              Num_of_ran = Int(round(N*0.005))
              ran=randperm(N)[1:Num_of_ran]
              qq=(1:Num_of_ran)/(Num_of_ran+1)
              temp1=floor.(cquantile.(Exponential(1/Infectious_parameter), qq))[randperm(Num_of_ran)]
              temp1[temp1.==0].=1
              Population[ran,Time_til_infected]=temp1  ## Time they will become infected and can pass on virus
              #Population[ran,]$Infected=0
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

```
                Population[ran,Susceptible].=1
                temp = floor.(cquantile.(Exponential(1/beta), qq))[randperm(Num_of_ran)]
                temp[temp.==0].=1
                Population[ran,time_of_next_inf] = Population[ran,Time_til_infected] + temp  ## Time they pass on virus to someone
                temp2 = floor.(cquantile.(Gamma(k,1/(k*sigma)), qq))[randperm(Num_of_ran)]
                Population[ran,time_of_recovery] = Population[ran,Time_til_infected] + temp2 # Time this person recovers when infected



                ## Create variables for tracking

                Susceptible_track       = zeros(Time)     ## vector created to track those susceptible in whole population
                Infected_track          = zeros(Time)     ## vector created to track those infected in whole population
                Recovered_track         = zeros(Time)     ## vector created to track those recovered in whole population
                Second_group_track      = zeros(Time)     ## vector created to track those recovered in vulnerable population
                Second_group_track_susc = zeros(Time)     ## vector created to track those susceptible in vulnerable population
                Second_group_track_Inf  = zeros(Time)     ## vector created to track those infected in vulnerable population

                ## Begin simulation

                time_to_start_isolation=Time
                for t in 1:Time
                ##Isolation code. Only initialise isolation when proportion of suceptible people = quantile

                if (sum(Population[:,Susceptible])/N < quantile) && (t<time_to_start_isolation)
                time_to_start_isolation=t
                end

                ## Index the current people who have the chance to infect others.

                Current = Population[(Population[:,Time_til_infected].<t) .& (Population[:,Time_til_infected].>Int(0)) .& (Population[:,Recovered].==Int(0)),1]

                for i in Current
                Population[i,Infected] = 1                ## Person now becomes infected
                Population[i,Susceptible] = 0             ## Remove susceptible flag
                Population[i,previous_infection] = 1      ## Initialise previous infection flag

                ## If it is time to infect and not time to recover proceed

                if (Population[i,time_of_next_inf]==t & t<Population[i,time_of_recovery] )

                if ( Population[i,Group]==1  )

                ## Before isolation. We take alpha and not alpha prime
                if (t<time_to_start_isolation)
                X = ifelse(rand()<alpha,2,1)   ## Sample from which group to infect
                ## Randomly select one person from that group
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

```
                Y = Int(ifelse(X==1,ceil(G1_split*rand()),G1_split+ceil(G2_split*rand())))
                ## Check if person has virus previously. Otherwise give virus
                if ( Population[Y,previous_infection]==0 & Population[Y,Time_til_infected]==0 )
                temp = floor(rand(Exponential(1/beta)))
                Population[Y,Time_til_infected]= t+ floor(rand(Exponential(1/Infectious_parameter)))
                Population[Y,time_of_next_inf] = Population[Y,Time_til_infected]+ifelse(temp==0,1,temp)
                Population[Y,time_of_recovery] = Population[Y,Time_til_infected]+floor(rand(Gamma(k,1/(k*sigma))))
                end
                ## isolation initialised. We take alpha prime here
                else
                X = ifelse(rand()<alpha_prime,2,1)   ## Sample from which group to infect
                ## Randomly select one person from that group
                Y = Int(ifelse(X==1,ceil(G1_split*rand()),G1_split+ceil(G2_split*rand())))
                ## Check if person has virus previously. Otherwise give virus
                if ( Population[Y,previous_infection]==0& Population[Y,Time_til_infected]==0 )
                temp = floor(rand(Exponential(1/beta1)))
                Population[Y,Time_til_infected]= t+ floor(rand(Exponential(1/Infectious_parameter)))
                Population[Y,time_of_next_inf] = Population[Y,Time_til_infected]+ifelse(temp==0,1,temp)
                Population[Y,time_of_recovery] = Population[Y,Time_til_infected]+floor(rand(Gamma(k,1/(k*sigma))))
                end
                end


                ## Repeat the same above for Group 1 but for Group 2 (Vulnerable group)

                elseif ( Population[i,Group]==2  )

                if (t<time_to_start_isolation)
                X = ifelse(rand()<alpha,2,1)
                Y = Int(ifelse(X==1,ceil(G1_split*rand()),G1_split+ceil(G2_split*rand())))
                if ( Population[Y,previous_infection]==0& Population[Y,Time_til_infected]==0  )
                temp = floor(rand(Exponential(1/beta)))
                Population[Y,Time_til_infected]= t+ floor(rand(Exponential(1/Infectious_parameter)))
                Population[Y,time_of_next_inf] = Population[Y,Time_til_infected]+ifelse(temp==0,1,temp)
                Population[Y,time_of_recovery] = Population[Y,Time_til_infected]+floor(rand(Gamma(k,1/(k*sigma))))
                end
                else
                X = ifelse(rand()<alpha_prime,2,1)
                Y = Int(ifelse(X==1,ceil(G1_split*rand()),G1_split+ceil(G2_split*rand())))
                if ( Population[Y,previous_infection]==0 & Population[Y,Time_til_infected]==0  )
                temp = floor(rand(Exponential(1/beta1)))
                Population[Y,Time_til_infected]= t+ floor(rand(Exponential(1/Infectious_parameter)))
                Population[Y,time_of_next_inf] = Population[Y,Time_til_infected]+ifelse(temp==0,1,temp)
                Population[Y,time_of_recovery] = Population[Y,Time_til_infected]+floor(rand(Gamma(k,1/(k*sigma))))
                end
                end
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

```
                end


                ## If it is not time for the person infected to recover, find next time the person will transmit
                if (t<time_to_start_isolation)
                temp = floor(rand(Exponential(1/beta)))
                Population[i,time_of_next_inf]=Population[i,time_of_next_inf]+ifelse(temp==0,1,temp)
                else
                temp = floor(rand(Exponential(1/beta1)))
                Population[i,time_of_next_inf]=Population[i,time_of_next_inf]+ifelse(temp==0,1,temp)

                end

                ## If it is time to recover, set recovery flag =1 and remove infected flag.

                elseif ( Population[i,time_of_recovery]<t )
                Population[i,Recovered]=1
                Population[i,Infected]=0
                end
                end # for i in Current

                ## Update tracking variables

                index = t

                Susceptible_track[index]   = sum(Population[:,Susceptible])/N ## Keep track of susceptible rate
                Recovered_track[index]     = sum(Population[:,Recovered])/N ## Keep track of recovered rate
                Infected_track[index]      = sum(Population[:,Infected])/N ## Keep track of infected rate
                Second_group_track[index]  = sum(Population[G1_split+1:N,Recovered])/G2_split   ## Keep track of recovered rate for vulnerable
                Second_group_track_susc[index] = sum(Population[G1_split+1:N,Susceptible])/G2_split ## Keep track of susceptible rate for vulnerable
                Second_group_track_Inf[index]  = sum(Population[G1_split+1:N,Infected])/G2_split  ## Keep track of Infected rate for vulnerable

                #global t=t+1
                end # for t

                Output     =  zeros(Time,8)
                Output[:,1] = Susceptible_track
                Output[:,2] = Recovered_track
                Output[:,3] = Infected_track
                Output[:,4] = Second_group_track
                Output[:,5] = Second_group_track_susc
                Output[:,6] = Second_group_track_Inf
                Output[:,7] .= time_to_start_isolation
                Output
                end")

# Set the parameters for the study
R=2.5
R_prime=2.5
N=1000000
c=1
quantile=0.8
alpha=0.2
sigma=1/21

#########################
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

```
Scenario_1  =  julia_call("Main_procedure",Total_time,R,R_prime,N,c,quantile,
    alpha,sigma)
## Change c for different scenarios
c=0.25
Scenario_2  =  julia_call("Main_procedure",Total_time,R,R_prime,N,c,quantile,
    alpha,sigma)
#########################
Scenario_1=t(Scenario_1)
Scenario_2=t(Scenario_2)

##  Plots for scenario 1
plot(seq(0,Total_time-1,1),Scenario_1[1,],type='l',col='red',ylim=c(0,1),lwd=3,
    lty=1,main='',xlab='Days',ylab='',cex.axis=2,cex.lab=2,cex.main=1.5)
lines(seq(0,Total_time-1,1),Scenario_1[2,],type='l',col='blue',lwd=3 ,lty=1  )
lines(seq(0,Total_time-1,1),Scenario_1[3,],type='l',col='green',lwd=3 ,lty=1  )
abline(v= Scenario_1[7,],lwd=2)

##  Lines for scenario 2
lines(seq(0,Total_time-1,1),Scenario_2[1,],type='l',col='red',ylim=c(0,1),lwd=3,
    lty=2,main='',xlab='Days',ylab='',cex.axis=2,cex.lab=2,cex.main=1.5)
lines(seq(0,Total_time-1,1),Scenario_2[2,],type='l',col='blue',lwd=3,lty=2)
lines(seq(0,Total_time-1,1),Scenario_2[3,],type='l',col='green',lwd=3,lty=2)
lines(seq(0,Total_time-1,1),Scenario_2[4,],type='l',col='black',lwd=3,lty=1  )
lines(seq(0,Total_time-1,1),Scenario_2[5,],type='l',col='orange',lwd=3,lty=1  )
lines(seq(0,Total_time-1,1),Scenario_2[6,],type='l',col='purple',lwd=3,lty=1  )
```

References

[1] Daley D.J., Gani J. Epidemic modelling: an introduction. Cambridge University Press, 2001.

[2] Hethcote H. W. The mathematics of infectious diseases. SIAM review: vol. 42, No. 4, pp. 599–653., 2000.

[3] Ferguson M.N. et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. https://www.imperial.ac.uk/media/imperial-college/\medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf, 2020.

[4] "https://bylinetimes.com/2020/03/23/covid-19-special-investigation-part-one-the-\politicised-science-that-nudged-the-johnson-government-to-safeguard-the-economy-over-british-lives/".

[5] "https://www.spiegel.de/wissenschaft/medizin/coronavirus-die-zahlen-sind-vollkommen-unzuverlaessig-a-7535b78f-ad68-4fa9-9533-06a224cc9250".

[6] Shen C., Taleb N. N., Bar-Yam Y. Review of ferguson et al "impact of non-pharmaceutical interventions...". "https://necsi.edu/review-of-ferguson-et-al-impact-of-non-pharmaceutical-interventions".

[7] Ghafari M., Moritz Kraemer M., Thompson C., Simmonds P., Klenerman P., Gupta S. Lourenco J., Paton R. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. Doi: "https://doi.org/10.1101/2020.03.24.20042291", 2020.

[8] Tang B., Bragazzi N.L., Li Q., Tang S., Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infectious Disease Modelling, vol. 5, pp. 248–255.", 2020.

[9] Kucharski A., Russell T., Diamond C., Liu Y. Analysis and projections of transmission dynamics of nCoV in Wuhan. CMMID repository, vol. 2.", 2020.

[10] "www.womenshealthmag.com/health/a31284395/how-long-does-coronavirus-last/".

[11] "https://www.npr.org/sections/goatsandsoda/2020/03/27/821958435/\why-death-rates-from-coronavirus-can-be-deceiving?t=1585335235282".

medRxiv preprint doi: https://doi.org/10.1101/2020.04.04.20053462; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[12] Coronavirus: elderly in lockdown and children in school help Sweden pursue herd immunity, The Sunday Times. "www.thetimes.co.uk/edition/world/coronavirus-elderly-in-lockdown-and-children-in-school-help-sweden-pursue-herd-immunity-r705m76dd", 2020.

[13] "https://www.haaretz.com/israel-news/.premium.MAGAZINE-israeli-expert-trump-is-right-about-covid-19-who-is-wrong-1.8691031".

[14] "https://www.medrxiv.org/content/10.1101/2020.02.12.20022566v1".

[15] Cirillo P., Taleb N. N. Tail Risk of Contagious Diseases. "https://www.academia.edu/42307438/Tail_Risk_of_Contagious_Diseases", 2020.

[16] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.

[17] "https://www.spiegel.de/wissenschaft/medizin/coronavirus-ist-eine-kontaktsperre-nur-fuer-risikogruppen-eine-alternative-a-0f25ccea-780b-47ce-a06f-f76e98e6924d".

[18] "www.statista.com/statistics/281174/uk-population-by-age/".

[19] Zhigljavsky A. Theory of global random search. Kluwer Academic Publishers., 1991.

[20] Zhigljavsky A., Zilinskas A. Stochastic global optimization. Springer., 2007.

[21] Moskvina V., Zhigljavsky A. An algorithm based on singular spectrum analysis for change-point detection. Communications in Statistics – Simulation and Computation, 32(2), 319-352., 2003.

[22] Tartakovsky A., Nikiforov I., Basseville M. Sequential analysis: Hypothesis testing and change-point detection. CRC Press., 2014.

[23] Myerson R. B. Game theory. Harvard university press., 2013.

[24] Watanabe S., Nair M. G., Rajeev B. Lectures on stochastic differential equations and Malliavin calculus. Springer., 1984.

[25] Anh V. V., Leonenko N. N. Renormalization and homogenization of fractional diffusion equations with random data. Probability Theory and Related Fields, 124(3), 381-408., 2002.