Discrete SIRIR modelling using empirical infection data shows that SARS-CoV-2 infection provides short-term immunity

Andrew McMahon[1,*] and Nicole C. Robb[1]

[1]Biological Physics, Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, OX1 3PU, United Kingdom

[*]To whom correspondence should be addressed: andrew.mcmahon@physics.ox.ac.uk

Abstract

The novel coronavirus SARS-CoV-2, which causes the COVID-19 disease, is now a global pandemic. Since December 2019, it has infected millions of people, caused the deaths of hundreds of thousands, and resulted in incalculable social and economic damage. Understanding the infectivity and transmission dynamics of the virus is essential for understanding how best to reduce mortality whilst ensuring minimal social restrictions to the lives of the general population. Anecdotal evidence is available, but detailed studies have not yet revealed whether infection with the virus results in immunity. In this work, we have extended the generic SIR framework to analyse empirical infection and fatality data from different regions to investigate the reinfection frequency of the disease. Our model predicts that cases of reinfection should have been observed by now if primary SARS-CoV-2 infection did not protect from subsequent exposure in the short term, however, no such cases have been documented. This work, therefore, provides a useful insight for serological testing strategies, lockdown easing and vaccine design.

Introduction

The novel coronavirus SARS-CoV-2 is thought to have originated in China in late 2019, and has since spread globally, resulting in the COVID-19 pandemic. The virus is a respiratory pathogen that causes mild symptoms in the majority of cases, but can progress to Acute Respiratory Distress Syndrome (ARDS) in a small number of people, which can result in death [1]. To date, the virus has resulted in over six million confirmed infections, almost 400,000 deaths and caused huge social and economic damage. Over 90% of countries have implemented travel restrictions [2] and countless businesses have been closed.

The spike protein of the coronavirus is thought to be the primary antigenic target against which neutralizing antibodies are produced [3] [4] [5]. Longitudinal profiles of antibody responses in recovered SARS-CoV-1 patients showed that IgG levels were maintained in >90% of patients for 2 years, however this dropped to approximately 50% of the convalescent population after 3 years [6]. While it is too soon to predict the long term serological response of SARS-CoV-2, it has been shown in a study of 285 individuals presenting with COVID-19 that 100% tested positive for IgG [7], and a study of hospital staff in France also supported the use of serologic testing to diagnose those who have recovered from SARS-CoV-2 [8]. Rhesus macaque monkeys that had recovered from a primary SARS-CoV-2 infection showed no recurrence of COVID-19 symptoms upon being re-challenged with the virus [9] [10], however, it is not clear whether this immunity will apply to the human population, or how long it may last for. It is therefore important to investigate whether infection provides immunity against further reinfection, as this can open up new quarantine easing strategies and decrease the further social and economic burden of the virus.

SIR (Susceptible, Infected, Recovered) modelling uses a set of differential equations to determine how the number of infected and recovered individuals changes over time given a specified rate of infection and recovery. It was first used in 1927 by Kermack and McKendrick [11] and has since been used to model epidemics from Acquired Immune Deficiency Syndrome (AIDS) [12] to SARS [13]. Variations of SIR modelling have been used during the COVID-19 pandemic to look at the varying burden on healthcare systems based on public health intervention [14], the absence of a stable disease-free equilibrium [15] and infection rate [16], as well as the eventual size of the overall pandemic [17]. An extension of the model has also been used to simulate the changing death rate as a function of the number of individuals infected, and it was found that an equilibrium point was reached where there are no further reinfections [18].

In this study, we have devised a simple dynamic model that uses empirical data taken from a compiled COVID-19 dataset [19] to investigate the reinfection frequency of the disease. By extending the generic SIR framework, we have separated the original infection from subsequent infections to produce our SIRIR (Susceptible, Infected, Recovered, Infected (two or more times), and Recovered (two or more times)) framework to investigate the susceptibility of a person to reinfection. The results of this analysis showed that a small number of cases classified as 'reinfections' should have occurred, however, no definitive cases of reinfection have been reported in the scientific literature to date. This suggests that primary SARS-CoV-2 infection is effective at preventing reinfection in the short term.

Results

Simulations of United Kingdom infection data suggest a low number of reinfections should have occurred

We initially set out to investigate the number of potential reinfection cases by modelling infection data from the United Kingdom [19]. We based our model on the compartmental SIR framework, but differentiated between initial and subsequent infections to produce a SIRIR (Susceptible, Infected, Recovered, Infected (two or more times), and Recovered (two or more times)) model (Figure 1 & Table 1).

When using this model, the rate of infection, recovery and fatality were assumed to be independent of how many previous infections a host had previously had. The number of infections, deaths and recoveries per day and the populations of the regions were taken from national statistics [19]. The number of susceptible persons at the beginning of the simulation, N, was taken to be the population of the region of interest [19] [20]. After all infections, recoveries and deaths for a day, the number of days into the simulation was increased by one, $t \rightarrow t+1$ up to t_{max}. The simulation was repeated 10,000 times to produce expectation values and standard deviations for the number of individuals classified as reinfections.

We initially ran the simulation for data in the United Kingdom [19] over the course of 106 days (from the first recorded case on the 1st of February until the 17th of May 2020 when the data was accessed). Figure 2 shows how the population of each state in the model changed over the course of a typical simulation. The number of susceptible individuals initially remained steady, until day 55, when there was a sharp decline due to the increase in primary infections (Figure 2A). The number of individuals infected just once started to increase steadily after day 40, and continued to do so throughout the simulation until lockdown effects enacted on day 53 (the 23rd of March 2020) took effect to control the number infected. The average recovery time used in the simulations was set as 28 days, as this is greater than the median recovery time suggested in the report of the WHO-China joint mission on coronavirus disease 2019 [21]. After the 28-day lag time, the individuals infected once started to recover, resulting in an increase in the Recovered (once) state through to the end of the simulation (Figure 2B). As the number of recovered individuals started to increase, so did the number of people infected for a second time, and similarly the number of people recovered for the second time started to increase after the 28 day recovery lag time (Figure 2C). The number of deaths started to rise from day 55 onwards, and fatalities continued to increase through to the end of the simulation (Figure 2D).

By pooling the number of cases in the Infected (two or more times), Recovered (two or more times) and those deceased from the Infected (two or more times) states at the end of the simulation, we calculated an estimate of the number of reinfections that would be expected to occur. This number represents the total population that had passed through the infected (two or more times) state by the end of the simulation. In the United Kingdom, the number of expected reinfections is low (70±8), particularly as a percentage of the total number of infections (Table 2). However, it is greater than zero, suggesting that reinfections should have already been seen. As no definitive cases of reinfection have been reported to date, this suggests that initial SARS-CoV-2 infection does provide immunity over the time period of our simulation.

Simulations of infection data in other regions show a similar trend

The simulations were repeated with data from Australia [19], Italy [19], New York City [22] [20], Singapore [19], Switzerland [19] and the United States of America (USA) [19]. In each case the model was run from the day of the first confirmed infection in the location to the 17th of May 2020, when the data was accessed. The mean number of expected reinfections in each region or country for the 10,000 simulations that were run are shown in Table 2.

In Australia, the number of confirmed SARS-CoV-2 infections to date has been relatively low [19], possibly due to early social distancing measures, the closing of international borders and mass testing and tracing measures. The number of modelled reinfections (0.1±0.3; Table 2) reflects this, and so even without immunity from infection no reinfections would be expected to occur. Similarly, in Switzerland and Singapore, very low numbers of reinfections were predicted by the model (6.2±2.5 and 6±2 respectively; Table 2). It is possible that these very low numbers of reinfection cases could have been missed due to misdiagnosis or lack of follow up testing. We therefore applied our model to data from Germany [19], Italy [19], New York City [20] [22] and the USA as a whole [19], which have recorded far higher numbers of SARS-CoV-2 infections (174355, 224760, 189027 and 1467884 respectively, when the data was accessed). The number of reinfection cases predicted for these countries was 30±6, 89±9, 335±18 and 635±25 for Germany, Italy, New York and the USA respectively (Table 2). We

conclude that it is therefore very unlikely that all of these predicted cases, if true, were missed due to misdiagnosis or lack of testing. The actual number of cases of reinfection that have been reported in any of these regions or countries to date is zero, suggesting that worldwide, primary SARS-CoV-2 infection provides short term immunity.

Comparison of infection and hospitalisation data in New York

Next, we repeated our simulation for New York City with the total number of infections replaced by the number of hospitalisations [23]. We hypothesised that hospitalised patients would be more likely to be clinically followed up and reinfections would be documented if they occurred. When we ran the simulation with an input of the total number of infections, the number of secondary infections continued to increase until day 75 (just before the end of the simulation on day 77), when the numbers appear to peak (Figure 3A). This was followed by an increase in the number of secondary recoveries after the 28-day recovery lag time. In comparison, the hospitalisation data for New York showed no secondary recoveries as the reinfections occurred later into the simulation (Figure 3B). The total number of predicted reinfections from the New York hospitalised data was 12±4 (Table 2), however, there are no documented cases of rehospitalisations to date. It is unlikely that these cases would be missed, or that symptoms would be wrongly assigned as a different disease as people would be processed and fully tested on admission to hospital.

Inclusion of recovery data suggests that predicted reinfections are underestimated

Recovery data is sparse or not available for most regions, likely due to lack of follow-up testing. In Germany, however, some recovery data was available [24]. We therefore compared the results of our simulation for Germany with and without the recovery data as an input. Without recovery data, the model used a 28-day lag before recoveries started, meaning very few secondary recoveries took place (Figure 4A). In contrast, when recovery data was included secondary recoveries began very soon after reinfections had taken place, usually the next day, and by the end of the simulation, almost the entire infected population had recovered (Figure 4B). Because secondary recoveries occurred faster, there were 76 more reinfections with the reinfection data than with the modelled data (Table 2). This suggests that we have in fact underestimated the number of predicted reinfections in our model; as no documented reinfection cases have been reported this strongly supports our conclusion that initial SARS-CoV-2 infection is effective at preventing reinfection in the short term.

The 28-day lag time used for the modelled recovery data ensured that we underestimated the recovery rate, and so also the rate of reinfection. To investigate a more life-like recovery rate, the United Kingdom simulations were repeated again using the modelled recovery data, while shortening the lag time for recovery. As expected, we found that the rate of reinfection was increased as the lag time was decreased from 28 days through to 7 days, as there was a larger population that recovered from a primary infection. With a 7-day lag time, the number of people in the infected (2 or more times) state peaked at around day 100 of the simulation (Figure S1). The total number of people reinfected throughout the simulation increased with a decreasing lag time, with 70±8, 121±11, 187±14 and 267±16 reinfections for 28-day, 21-day, 14-day and 7-day recovery lag times respectively. With a median recovery time of one to two weeks for mild cases [21], the 7-day or 14-day lag times represent more realistic figures for the recovery rate. It is therefore extremely unlikely that the large number of reinfections that these lag times suggest would have been missed, and so we conclude that initial infection results in immunity against SARS-CoV-2.

Discussion

In this work we have presented a modelling strategy to determine whether SARS-CoV-2 reinfections can occur. We modelled actual infection and fatality data from different regions and countries around the world and found that all regions investigated, with the exception of Australia, should have recorded cases of reinfections if primary infection with SARS-CoV-2 did not provide some level of immunity. Australia may be an exception as the country was relatively quick to adopt early social distancing measures such as mass testing and contact tracing, as well as preventing international travel from high risk countries [25]. We also found that rehospitalisation cases should have been seen amongst hospitalised cases in New York City. To date, however, no reinfections have verifiably been recorded anywhere in the world. A report from South Korea suggested that 116 patients recovered from COVID-19 had tested positive by RT-PCR for the virus again [26], however, this has since been explained as the 'false-positive' detection of remnants of viral RNA rather than reactivation or reinfection. The lack of documented reinfections suggests that short-term immunity to the virus is produced by an initial infection, however, our model cannot predict whether this immunity will last over longer time scales.

Our results are supported by a number of animal challenge studies. A study in rhesus macaques showed that, following initial viral clearance, the monkeys showed a reduction in their median viral load in comparison with primary infection when rechallenged with SARS-CoV-2 [10]. Similarly, Ryan et al. demonstrated that rechallenged ferrets were fully protected from acute lung pathology [27]. Finally, an adenovirus-vector vaccine tested on rhesus macaques elicited a humoral and cellular response that, on challenge with the virus, proved to significantly reduce the viral load in bronchoalveolar lavage fluid and respiratory tract tissue [28].

A report from the WHO-China joint mission on Coronavirus disease estimated the recovery time for SARS-CoV-2 infection as 2 weeks for mild cases and 3-6 weeks for severe or critical disease [21]; based on this we used a long (28 day) recovery lag time in the modelled data. Comparison with real-world recovery data from Germany suggested that the actual recovery time may be significantly shorter, giving rise to an underestimation of the reinfection rate in our modelled data. This was supported by an increase in the number of predicted reinfections in the United Kingdom simulations when we used a shorter recovery lag time of 7, 14 or 21 days. In addition, there were no allowances in our model for transmission being localised to regions smaller than a nation or state; the daily infection data was likely to be only a fraction of the total number of infections due to asymptomatic or mild infections not being recorded, and infections were recorded on the date of testing, not the actual date of infection. We also note that significant differences in testing, reporting and shielding of the vulnerable exist between the different regions in this study. In every region, we expect that the impact on our simulation would be to underestimate the number of reinfections, though these differences mean direct comparisons between countries are not valid. Taken together, this suggests that the actual reinfection rate would be significantly higher than that predicted by our model if there was no immunity conferred by prior infection.

The results documented here provide strong evidence, based on real data, to suggest that that there is at least short-term immunity conferred by an initial infection of SARS-CoV-2. This has implications for serological testing strategies, lockdown easing timescales and vaccine development. Our modelling strategy can also be extended to understand the reinfection dynamics of future pandemics.

Methods

We have developed an extension to the discrete SIR model and used national infection and mortality data [19] [20] [22] [23] to investigate the reinfection dynamics of SARS-CoV-2. In our model, it was assumed that the infection rate was unchanged by prior infection, and so a comparison of real reinfections against modelled reinfections provides an insight into the true reinfection dynamics.

Assumptions

A number of assumptions have been made. Where possible, they have been made so that the number of reinfections is underestimated. These assumptions are:

1. There is a large lag time for a recovery to take place (28 days).
2. The model does not consider social distancing or shielding and so assigns an equal probability of an infection to anyone.
3. Not all infections have been recorded due to lack of testing, misdiagnosis or asymptomatic infection.
4. Infections and recoveries are not necessarily recorded on the date that they first occurred.
5. There is no emigration out of, or immigration into, a population of interest.
6. Hospitalisations are limited to people previously hospitalised.
7. A homogeneous population density with no societal structure (e.g. equal residents per household).

Assignment of probabilities is given by the size of the state and the total number of infected individuals

The probability of an infection is determined by the ratio of the population of the state being infected to the total number of uninfected individuals, $N_t^{uninfected}$. The probability of any recovery or death is given by the ratio of the population recovering/dying to the total number of infected states, $N_t^{infected}$. For example, for the probability of an infection from the Susceptible state S_t to the infected (first time) state I_t is given by:

$$p(S_{t-1} \rightarrow I_t | Infection) = \frac{\beta_t S_{t-1}(I_{t-1} + I'_{t-1})}{\beta_t S_{t-1}(I_{t-1} + I'_{t-1}) + \beta_t R_{t-1}(I_{t-1} + I'_{t-1}) + \beta_t R'_{t-1}(I_{t-1} + I'_{t-1})}$$
$$= \frac{S_{t-1}}{S_{t-1} + R_{t-1} + R'_{t-1}} = \frac{S_{t-1}}{N_{t-1}^{uninfected}} \quad (1)$$

where $\beta_t S_{t-1}(I_{t-1} + I'_{t-1})$ is the number infected from the Susceptible state, S_{t-1}, to the Infected (once) state I_t; $\beta_t R_{t-1}(I_{t-1} + I'_{t-1})$ is the number infected from the Recovered (once) state, R_{t-1}, to the Infected (two or more times) state I'_t and $\beta_t R'_{t-1}(I_{t-1} + I'_{t-1})$ is the number infected from the Recovered (two or more times state, R'_{t-1}, to the Infected (two or more times) state I'_t.

The model equations

The equations relating how each of the states (Susceptible, Infected (once), Recovered (once), Infected (two or more times), Recovered (two or more times) and Deceased respectively) changed for each iteration of the simulation were:

$$S_t = S_{t-1} - \beta_t S_{t-1}(I_{t-1} + I'_{t-1}) \quad (2)$$

$$I_t = I_{t-1} + \beta_t S_{t-1}(I_{t-1} + I'_{t-1}) - \gamma_t I_{t-1} - m_t I_{t-1} \quad (3)$$

$$R_t = R_{t-1} + \gamma_t I_{t-1} - \beta_t R_{t-1}(I_{t-1} + I'_{t-1}) \quad (4)$$

$$I'_t = I'_{t-1} + \beta_t R_{t-1}(I_{t-1} + I'_{t-1}) - \gamma_t I'_{t-1} + \beta_t R'_{t-1}(I_{t-1} + I'_{t-1}) - m_t I'_{t-1} \quad (5)$$

$$R'_t = R'_{t-1} + \gamma_t I'_{t-1} - \beta_t R'_{t-1}(I_{t-1} + I'_{t-1}) \quad (6)$$

$$D_t = D_{t-1} + m_t I_{t-1} + m_t I'_{t-1} \quad (7)$$

where the symbols are defined in Table 1 and the processes are shown in Figure 1.

The uniform distribution used for attributing infections/recoveries/deaths

We have used the uniform distribution with range 1 to R, where R is a real integer, to determine states involved in infections, recoveries and deaths. The probability for every number in the range is equal and given by

$$p(x, R) = \frac{1}{R-1}, 1 \leq x \leq R. \quad (8)$$

For sampling, we have used the cumulative distribution function:

$$P(x, R) = \int_1^x p(x, R)\, dx = \frac{x-1}{R-1}, 1 \leq x \leq R \quad (9)$$

Assigning infections

The patients that were infected each day were chosen at random from the 3 states that are not infected with the virus (weighted by the number of persons in each state),

- $n_t^{infected}$ random numbers, x, are taken from the distribution $p(x, N_{t-1}^{uninfected})$
- The infections are assigned to states going through all the random numbers x_i:

$$\beta_t S_{t-1}(I_{t-1} + I'_{t-1}) \rightarrow \beta_t S_{t-1}(I_{t-1} + I'_{t-1}) + 1 \text{ if } P(x_i, N_{t-1}^{uninfected}) \leq P(S_{t-1}, N_{t-1}^{uninfected}) \quad (10)$$

$$\beta_t R_{t-1}(I_{t-1} + I'_{t-1}) \rightarrow \beta_t R_{t-1}(I_{t-1} + I'_{t-1}) + 1 \text{ if } P(x_i, N_{t-1}^{uninfected}) \leq P(S_{t-1} + R_{t-1}, N_{t-1}^{uninfected}) \quad (11)$$

$$\beta_t R'_{t-1}(I_{t-1} + I'_{t-1}) \rightarrow \beta_t R'_{t-1}(I_{t-1} + I'_{t-1}) + 1 \text{ if } P(x_i, N_{t-1}^{uninfected}) > P(S_{t-1} + R_{t-1}, N_{t-1}^{uninfected}) \quad (12)$$

where $\beta_t S_{t-1}(I_{t-1} + I'_{t-1})$ is the number of people infected from the Susceptible state, S_{t-1}, to the Infected (once) state I_t; $\beta_t R_{t-1}(I_{t-1} + I'_{t-1})$ is the number infected from the Recovered (once) state, R_{t-1}, to the Infected (two or more times) state I'_t; $\beta_t R'_{t-1}(I_{t-1} + I'_{t-1})$ is the number infected from the Recovered (two or more times state, R'_{t-1}, to the Infected (two or more times) state I'_t and $N_{t-1}^{uninfected}$ is the total number of people on day t-1 who are not in one of the infected states.

The number of infections each day, $n_t^{infected}$, is taken from the real number of infections that occurred in the region of interest on each day.

Assigning recoveries with recovery data

The patients that recover each day were chosen at random from the two states that were infected with the virus (weighted by the number of persons in each state).

- $n_t^{recovered}$ random numbers, x, are taken from the distribution $p(x, N_{t-1}^{infected})$
- The recoveries are assigned to states going through all the random numbers x_i:

$$\gamma_t I_{t-1} \rightarrow \gamma_t I_{t-1} + 1 \text{ if } P(x_i, N_{t-1}^{infected}) \leq P(I_{t-1}, N_{t-1}^{infected}) \quad (13)$$

$$\gamma_t I'_{t-1} \rightarrow \gamma_t I'_{t-1} + 1 \text{ if } P(x_i, N_{t-1}^{infected}) > P(I_{t-1}, N_{t-1}^{infected}) \quad (14)$$

where $\gamma_t I_{t-1}$ is the number recovering from the Infected (once) state, I_{t-1}, to the Recovered (once) state R_t; $\gamma_t I'_{t-1}$ is the number recovering from the Infected (two or more times) state, I'_{t-1}, to the Recovered

(two or more times) state I'_t and $N_{t-1}^{infected}$ is the total number of people on day t-1 who are in one of the infected states.

The number of infections each day, $n_t^{recovered}$, is taken from the real number of recoveries that occurred in the region of interest on each day.

Assigning recoveries without recovery data

The patients that recover each day were chosen at random from the two states that were infected with the virus (weighted by the number of persons infected into each state a number of days earlier, $t_{recovery}$).

- $n_t^{recovered}$ random numbers, x, are taken from the distribution $p(x, n_{t-t_{recovery}}^{infected})$
- The recoveries are assigned to states going through all the random numbers x_i:

$$\gamma_t I_{t-1} \rightarrow \gamma_t I_{t-1} + 1 \text{ if } P(x_i, n_{t-t_{recovery}}^{infected}) \leq P(i_{t-t_{recovery}}^{first\ time}, n_{t-t_{recovery}}^{infected}) \quad (15)$$

$$\gamma_t I'_{t-1} \rightarrow \gamma_t I'_{t-1} + 1 \text{ if } P(x_i, n_{t-t_{recovery}}^{infected}) > P(i_{t-t_{recovery}}^{first\ time}, n_{t-t_{recovery}}^{infected}) \quad (16)$$

where $\gamma_t I_{t-1}$ is the number recovering from the Infected (once) state, I_{t-1}, to the Recovered (once) state R_t; $\gamma_t I'_{t-1}$is the number recovering from the Infected (two or more times) state, I'_{t-1}, to the Recovered (two or more times) state I'_t ; $N_{t-1}^{infected}$ is the total number of people on day t-1 who are in one of the infected states and $i_{t-t_{recovery}}^{first\ time}$ is the number of people infected for the first time on day $t-t_{recovery}$.

The number of infections each day, $n_t^{recovered}$, is given by the number of persons infected a number of days earlier, $n_{t-t_{recovery}}^{infected}$ who would not be expected to have died

$$n_t^{recovered} = \left(1 - \frac{\sum_t n_t^{death}}{\sum_t n_t^{infected}}\right) n_{t-t_{recovery}}^{infected} \quad (17)$$

Where n_t^{deaths} is the number of deaths on day t and $n_t^{infected}$ is the number of people infected on day t.

The cycle between Infected (two or more times) and Recovered (two or more times) is permitted as we are only interested in reinfection rates.

Assigning deaths

The patients that die each day were chosen at random from the two states that were infected with the virus (weighted by the number of persons in each state).

- n_t^{deaths} random numbers, x, are taken from the distribution $p(x, N_{t-1}^{infected})$
- The recoveries are assigned to states going through all the random numbers x_i:

$$m_t I_{t-1} \rightarrow m_t I_{t-1} + 1 \text{ if } P(x_i, N_{t-1}^{infected}) \leq P(I_{t-1}, N_{t-1}^{infected}) \quad (18)$$

$$m_t I'_{t-1} \rightarrow m_t I'_{t-1} + 1 \text{ if } P(x_i, N_{t-1}^{infected}) > P(I_{t-1}, N_{t-1}^{infected}) \quad (19)$$

where $m_t I_{t-1}$ is the number dying from the Infected (once) state, I_{t-1}, to the Deceased state D_t; $m_t I'_{t-1}$ is the number dying from the Infected (two or more times) state, I'_{t-1}, to the Deceased state D_t (these are counted so the total number of individuals passing through the Infected (two or more times) state can be counted) and $N_{t-1}^{infected}$ is the total number of people on day t-1 who are in one of the infected states.

The number of deaths each day, n_t^{deaths}, is taken from the real number of deaths that occurred in the region of interest on each day.

The deaths from the Infected (two or more times) state are counted so that the sum of this, the Infected (two or more times) and the Recovered (two or more times) values give the total number of reinfections up to that point of the simulation.

Acknowledgements

We thank Dr. Barak Gilboa for critical reading of the manuscript. This work was supported by Royal Society Dorothy Hodgkin Research Fellowship DKR00620 and Research Grant for Research Fellows RGF\R1\180054 (to N.C.R.).

References

[1] World Health Organization, "Rolling updates on coronavirus disease (COVID-19)," [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen. [Accessed 20 May 2020].

[2] J. Asquith, "Coronavirus Travel Update - Over 90% Of People Live Under Travel Restrictions," Forbes, 2 April 2020. [Online]. Available: https://www.forbes.com/sites/jamesasquith/2020/04/02/april-update-coronavirus-travel-update-90-of-countries-have-travel-restrictions/. [Accessed 20 May 2020].

[3] Z. Y. Yang, W. P. Kong, Y. Huang and e. al., "A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice," *Nature,* vol. 428, pp. 561-4, 2004.

[4] Y. He, J. Li, S. Heck and e. al., "Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design," *J Virol,* vol. 80, 2006.

[5] U. J. Buchholz, A. Bukreyev, L. Yang and e. al., "Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity," *Proc Natl Acad Sci USA,* vol. 101, pp. 9804-9, 2004.

[6] L.-P. Wu, N.-C. Wang and Y.-H. e. a. Chang, "Duration of Antibody Responses after Severe Acute Respiratory Syndrome," *Emerging Infectious Diseases,* vol. 13, no. 10, pp. 1562-1564, 2007.

[7] Q.-X. Long, B.-Z. Liu and H.-J. e. a. Deng, "Antibody Responses to SARS-CoV-2 in Patients With COVID-19," *Nat Med,* Vols. https://doi.org/10.1038/s41591-020-0897-1, 2020.

[8] S. Fafi-Kremer, T. Bruel and Y. Madec, "Serologic responses to SARS-CoV-2 infection among hospital staff with mild disease in eastern Francer," *medRxiv,* 2020.

[9] L. Bao, W. Deng and G. Hong, "Reinfection could not occur in SARS-CoV-2 infected rhesus macaques," *bioRxiv,* 2020.

[10] A. Chandrashekar, J. Liu and A. J. e. a. Martinot, "SARS-CoV-2 infection protects against rechallenge in rhesus macaques," *Science,* 2020.

[11] W. O. Kermack, A. G. McKendrick and G. T. Walker, "A contribution to the mathematical theory of epidemics," *Proc. R. Soc. Lond. A,* vol. 115, pp. 700-721, 1997.

[12] A. O. Victor and H. K. Oduwole, “Evaluating the deterministic SERIUS model for disease control in an age-structured population,” *Global Sci J,* vol. 8, no. 3, 2020.

[13] T. W. Ng, G. Turinici and A. Danchin, “A double epidemic model for the SARS propagation,” *BMC Infect DIs.,* vol. 3, no. 19, 2003.

[14] W. K. Ming and C. J. Zhang, “Breaking down of healthcare system: mathematical modelling for controlling the novel coronavirus(2019-nCoV) outbreak in Wuhan, China,” *bioRxiv,* 2020.

[15] A. Victor, “Mathematical predictions for COVID-19 as a global pandemic,” *SSRN Journal,* 2020.

[16] I. Nesteruk, “Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China,” *medRxiv,* 2020.

[17] M. Batista, “Estimation of the final size of the coronavirus epidemic by the SIR model,” *ResearchGate,* 2020.

[18] A. V. Okheuse, “Estimation of the Probability of Reinfection With COVID-19 by theSusceptible-Exposed-Infectious-Removed-Undetectable-SusceptibleModel,” *JMIR Public Health and Surveillance,* 2020.

[19] H. Ritchie, M. Roser, E. Ortiz-Ospina and J. Hasell, “Coronavirus pandemic (COVID-19) - Country by country,” Our World in Data, [Online]. Available: https://ourworldindata.org/coronavirus-country-by-country. [Accessed 17 May 202].

[20] United States Census Bureau, “QuickFacts: Richmond County (Staten Island Borough), New York; New York County (Manhattan Borough), New York; Bronx County (Bronx Borough), New York; Kings County (Brooklyn Borough), New York; Queens County (Queens Borough), New York,” [Online]. Available: https://www.census.gov/quickfacts/fact/table/richmondcountystatenislandboroughnewyork,newyorkcountymanhattanboroughnewyork,bronxcountybronxboroughnewyork,kingscountybrooklynboroughnewyork,queenscountyqueensboroughnewyork/PST045219. [Accessed 17 May 2020].

[21] WHO, “Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19),” WHO, 2020.

[22] City of New York, “Covid-19: Data,” NYC.gov, [Online]. Available: https://www1.nyc.gov/site/doh/covid/covid-19-data.page. [Accessed 17 May 2020].

[23] Federal Office of Public Health FOPH, “New coronavirus: Situation in Switzerland,” [Online]. Available: https://www.bag.admin.ch/bag/en/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/situation-schweiz-und-international.html#-1199962081. [Accessed 17 May 2020].

[24] Trading Economics, “Germany Coronavirus Recovered,” [Online]. Available: https://tradingeconomics.com/germany/coronavirus-recovered. [Accessed 19 May 2020].

[25] L. Kelly, “Australia upbeat over slowing coronavirus spread, urges vigilance,” Reuters, 2020.

[26] S. Cha and J. Smith, “Explainer: South Korean findings suggest 'reinfected' coronavirus cases are false positives,” Reuters, 2020.

[27] K. A. Ryan, K. R. Bewley, S. A. Fotheringham and e. al., “Dose-dependent response to infection with SARS-CoV-2 in the ferret model: evidence of protection to re-challenge,” *bioRxiv,* 2020.

[28] N. van Doremalen, T. Lambe, A. Spencer and e. al., “ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques,” *bioRxiv,* 2020.

[29] R. Li, S. Pei and B. e. a. Chen, “Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2),” *Science,* vol. 368, no. 6490, pp. 489-493, 2020.

[30] KCDC, “Findings from investigation and analysis of re-positive cases,” 2020.

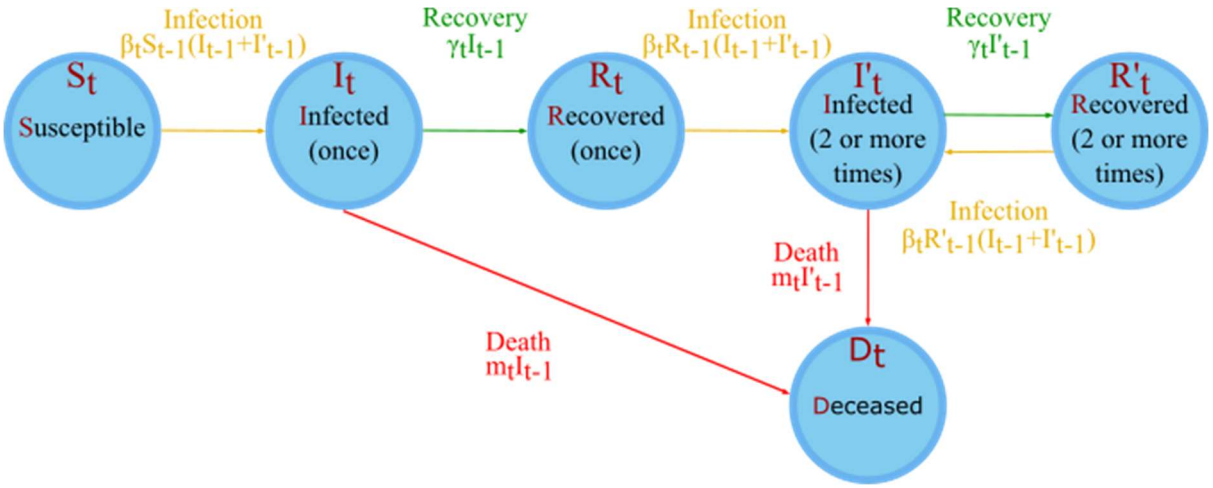

Figure 1. A simple representation of the model. S_t represents the number of persons susceptible to infection who have had no prior infections on day t; I_t is the number of people currently infected for the first time on day t; R_t is the number of people who have recovered once on day t; I'_t is the number of people who have been infected two or more times and are infected on day t; R'_t is the number of people who have recovered two or more times and are not infected on day t of the model and D_t is the number of deceased persons on day t of the model. Further symbols are defined in Table 1.

x_i	Random number
t	The number of days into the simulation. t=1 for the day of the first infection
$t_{recovery}$	The average number of days for recovery. Here, we have used $t_{recovery}$=28
t_{max}	The number of days over which the simulation is run
S_t	The number of susceptible on day t
I_t	The number of Infected (once) on day t
R_t	The number of Recovered (once) on day t
I'_t	The number of Infected (two or more times) on day t
R'_t	The number of Recovered (two or more times) on day t
D_t	The number of deceased from SARS-CoV-2 on day t
B_t	The infection rate on day t
γ	The recovery rate on day t
m_t	The death rate on day t
N	The total population in the model $N = S_t + I_t + R_t + I'_t + R'_t + Dt \quad \forall\, t$
$N_t^{infected}$	The number of infected individuals on day t $N_t^{infected} = I_t + I'_t$
$N_t^{uninfected}$	The number of uninfected individuals on day t $N_t^{uninfected} = S_t + R_t + R'_t$
$n_t^{infected}$	The number of new infections on day t $n_t^{infected} = \beta_t S_{t-1}(I_{t-1}+I'_{t-1}) + \beta_t R_{t-1}(I_{t-1}+I'_{t-1}) + \beta_t R'_{t-1}(I_{t-1}+I'_{t-1})$
$n_t^{recovered}$	The number of recovering individuals on day t $n_t^{recovered} = \gamma_t I_{t-1} + \gamma_t I'_{t-1}$
n_t^{deaths}	The number of deaths on day t $n_t^{deaths} = m_t I_{t-1} + m_t I'_{t-1}$
$i_t^{first\ time}$	The number of first time infections on day t. i.e. the number of $S_t \rightarrow I_t$ transitions on day t

Table 1. Definition of parameters in the Susceptible-Infected-Recovered-Infected (two or more times)-Recovered (two or more times) (SIRIR) model.

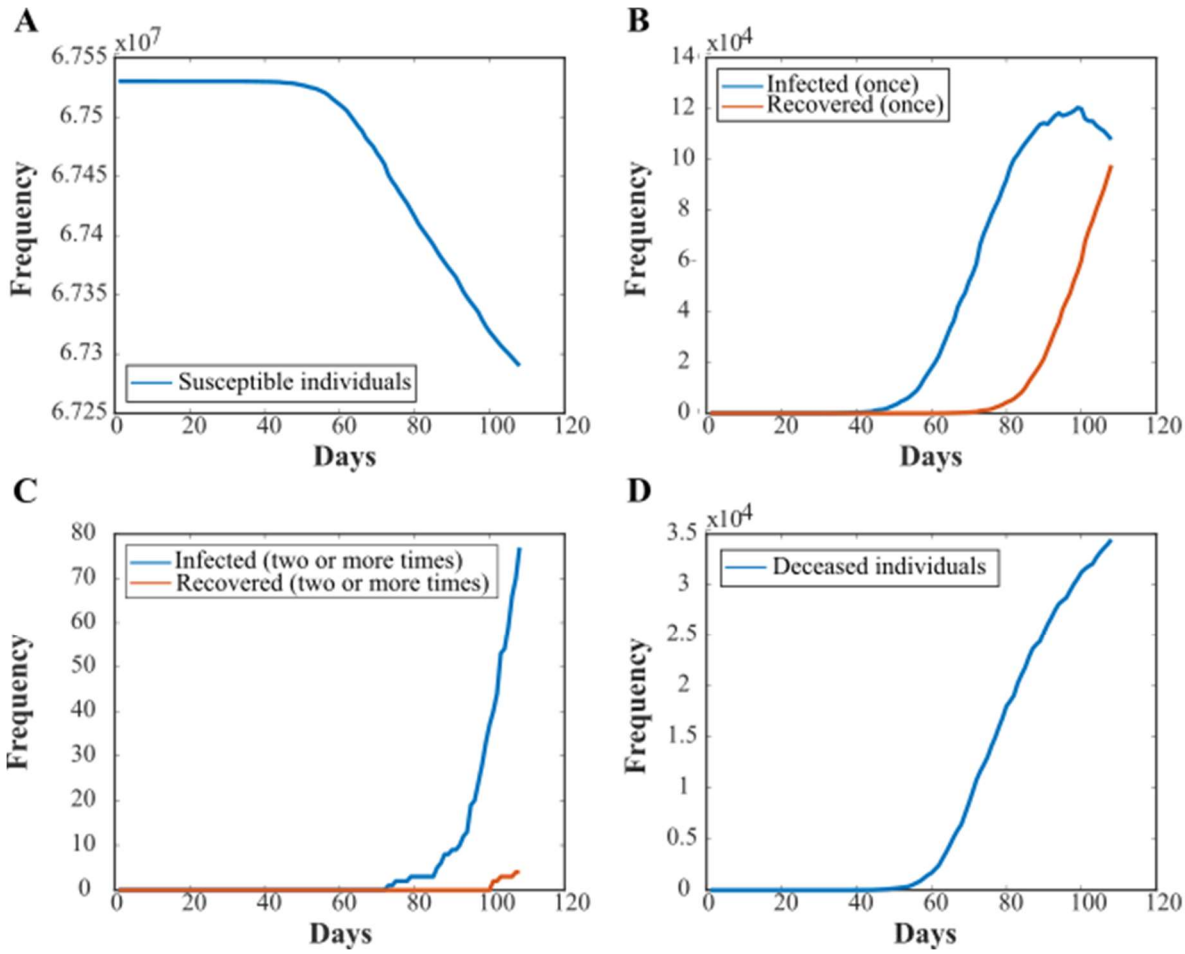

Figure 2. Plots of the populations of each state in the model over the course of a typical simulation, using infection data from the United Kingdom. A) An example plot of the susceptible population in the model over the course of 106 days (from the first recorded case on the 1st of February until the 17th of May 2020 when the data was accessed). B) An example plot of the populations that are infected for the first time or recovered from a single infection. C) An example plot of the populations of the simulation that have been reinfected and recovered from infection twice. D) An example plot of the number of deceased individuals through the course of the simulation.

Region	Mean number of reinfections	Standard deviation of number of reinfections	Total number of infections	Reinfections as a % of total infections
Australia	0.1	0.3	7036	0.001
Germany	30.0	6.0	174355	0.017
Germany (with recovery data)	106.0	10.0	174355	0.060
Italy	89.0	9.0	224760	0.040
New York City	335.0	18.0	189027	0.170
New York City (hospitalisations)	12.0	4.0	48462	0.006
Singapore	6.0	2.0	27356	0.022
Switzerland	6.2	2.5	30587	0.020
United Kingdom	70.0	8.0	240161	0.029
United States of America	635.0	25.0	1467884	0.040

Table 2. The number of predicted reinfections and their standard deviation in different locations worldwide as predicted from the model. Unless otherwise stated, these figures represent simulations using the total number of infections of each region and are modelled without the data on number of recoveries. S.D. standard deviation.

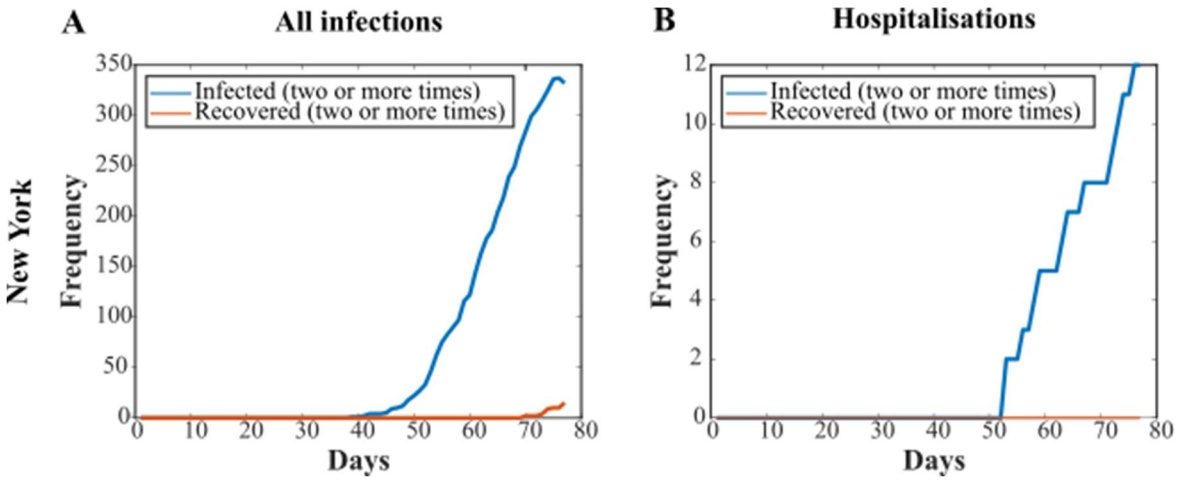

Figure 3. Comparison of total infections vs. hospitalisation data in New York City. Plots of the Infected (two or more times) and Recovered (two or more times) states for A) New York using all infection data B) New York using only the hospitalisations data.

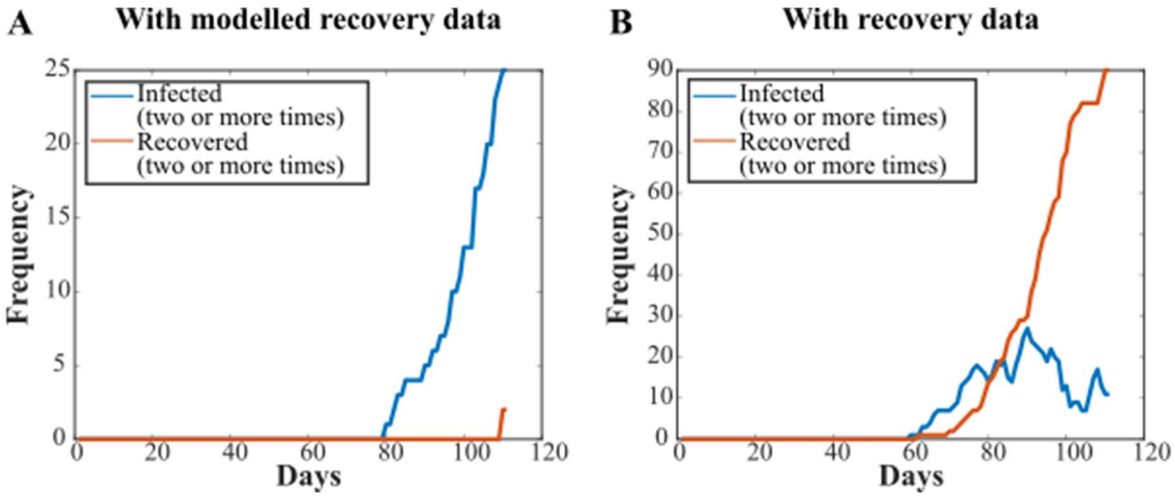

Figure 4. Use of actual recovery data from Germany suggests that the number of recovered individuals, and hence reinfections, are underestimated in our model. Plots of Infected (two or more times) and Recovered (two or more times) states with A) modelled recovery data and B) actual recovery data.