Dynamics of SARS-CoV-2 with Waning Immunity in the UK Population

Thomas Crellen1,*, Li Pi1, Emma L. Davis1, Timothy M. Pollington1,2, Tim C. D. Lucas1, Diepreye Ayabina1, Anna Borlase1, Jaspreet Toor1, Kiesha Prem3, Graham F. Medley3, Petra Klepac3,†, and T. Déirdre Hollingsworth1,†,*

1Big Data Institute, Old Road Campus, University of Oxford, Oxford, OX3 7LF, UK
2MathSys CDT, University of Warwick, Coventry, CV4 7AL, UK
3London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK
†Contributed equally
*Corresponding authors: thomas.crellen@bdi.ox.ac.uk & deirdre.hollingsworth@bdi.ox.ac.uk

July 24, 2020

Abstract

The dynamics of immunity are crucial to understanding the long-term patterns of the SARS-CoV-2 pandemic. While the duration and strength of immunity to SARS-CoV-2 is currently unknown, specific antibody titres to related coronaviruses SARS-CoV and MERS-CoV have been shown to wane in recovered individuals, and immunity to seasonal circulating coronaviruses is estimated to be shorter than one year. Using an age-structured, deterministic model, we explore different potential immunity dynamics using contact data from the UK population. In the scenario where immunity to SARS-CoV-2 lasts an average of three months for non-hospitalised individuals, a year for hospitalised individuals, and the effective reproduction number ($R_t$) after lockdown is 1.2 (our worst case scenario), we find that the secondary peak occurs in winter 2020 with a daily maximum of 409,000 infectious individuals; almost three-fold greater than in a scenario with permanent immunity. Our models suggests that longitudinal serological surveys to determine if immunity in the population is waning will be most informative when sampling takes place from the end of the lockdown until autumn 2020. After this period, the proportion of the population with antibodies to SARS-CoV-2 is expected to increase due to the secondary peak. Overall, our analysis presents considerations for policy makers on the longer term dynamics of SARS-CoV-2 in the UK and suggests that strategies designed to achieve herd immunity may lead to repeated waves of infection if immunity to re-infection is not permanent.

Introduction

As of 1st July 2020, SARS-CoV-2 has infected at least 10 million people worldwide and resulted in over 500,000 deaths [1, 2]. Following the initial outbreak from a live animal market in Wuhan, China [3], the United Kingdom (UK) has been among the countries most severely affected; reporting over 310,000 cases and 44,000 deaths, which is among the highest per-capita rates [2, 4]. Since 23rd March, nationwide non-pharmaceutical interventions (lockdown) have been in place to reduce social contacts by closing schools and shops; encouraging home working; and social distancing in public places. Similar measures have been in place in other European countries since late February 2020 with restrictions easing in France, Germany and Italy from May 2020. Within the European picture of disease control strategies, Sweden has been an outlier by placing fewer restrictions on social mixing while aiming to build up immunity in the population [5].

Following infection with the virus, hospitalised patients have an acute immune response where virus-specific IgM and IgG antibodies titres reach a maximum 15–21 and 22–27 days respectively after symptom onset [6, 7]. Antibodies raised in hospitalised patients and animal models against SARS-CoV-2 provide protection for at least several weeks following infection [8, 9], suggesting that immediate reinfection with the virus is unlikely. There is limited evidence that hospitalised patients with more severe symptoms show a greater antibody response [6, 9]. Asymptomatic individuals have a weaker IgG and specific antibody response to SARS-CoV-2 and are more likely to become seronegative following convalescence [10]. While the duration of immunity to SARS-CoV-2 is not currently known, antibody titres raised against related coronaviruses SARS-CoV and MERS-CoV have been shown to decay over time [11, 12]. Furthermore, immunity to seasonal circulating coronaviruses has been estimated to last for less than one year [13] and recovered individuals from coronavirus NL63 can become...
Dynamic epidemiological models play a major role in shaping the timing and intensity of interventions against SARS-CoV-2 in the UK and elsewhere [16]. Many models or simulations have assumed that infected individuals recover with permanent immunity [15, 17, 18]. In such models the epidemic reaches extinction after running out of infected individuals, although they do not preclude a second wave of infections after lockdown [19]. If immunity wanes over a period of time, or recovered individuals have only partial immunity to re-infection, this substantially alters the dynamics of the system [20]. In the absence of stochastic extinction and demography (births and deaths) in a population with equal mixing where; $R_0$ is the basic reproduction number; $\gamma$ is the average duration of infection; and $\omega$ is the reciprocal of the average duration of immunity; the endemic equilibrium proportion of infected in the population $I^*$ is driven by $(R_0 - 1) \omega / \gamma R_0$ and thus, in the absence of interventions, the infection persists indefinitely when $R_0 > 1$ [21].

In dynamic models which make the assumption of homogeneous mixing in the population, the ‘classic’ herd immunity threshold is given by $1 - 1 / R_0$. As $R_0$ for SARS-CoV-2 is generally estimated between 2.4–4 [22, 23, 24], this equates to 58–75% of the population requiring immunity to eventually halt the epidemic. Serological studies conducted in affected countries to-date have reported the proportion of the population with antibodies against SARS-CoV-2 to be much lower than this figure [22, 25]. However, when more realistic non-homogeneous mixing is considered, the observed herd immunity threshold is lower than the classical threshold [26]. Recent studies have considered this question for SARS-CoV-2 [27, 28], with Britton et al. noting that the disease-induced herd immunity threshold could be closer to 40% in an age-structured population, rather than the 60% classic herd immunity threshold when $R_0$ is 2.5 [28]. This phenomenon is driven by individuals that have more contacts, or greater susceptibility to the virus, getting infected earlier on and leaving the susceptible population; thus decelerating the growth of the epidemic.

Kissler et al. considered the dynamics of SARS-CoV-2 in the United States with seasonal forcing, homogeneous mixing and waning immunity that could be boosted by exposure to seasonal circulating betacoronaviruses [13]. Under these assumptions, the incidence of SARS-CoV-2 was predicted to rebound in winter months. Here we do not consider seasonality, but rather the dynamics of transmission in an age-structured population with different periods of waning immunity in the context of the UK emerging from lockdown.

We developed a discrete-time gamma delay-distributed (susceptible-exposed-infectious-recovered-susceptible; SEIRS) model, which incorporates current knowledge about the natural history of the virus and the UK population. Our model accounts for symptomatic and asymptomatic transmission, and heterogeneity in both daily contacts and infection susceptibility by age group. We consider different durations of immunity for hospitalised patients (or those with more severe symptoms) compared to non-hospitalised patients (those with less severe symptoms). We use this model to explore a range of scenarios in the UK population in the context of stringent non-pharmaceutical interventions (lockdown) followed by more limited interventions over a two year period from February 2020, and the impact of immunity duration on the longer term disease equilibrium.

Methods

Model structure

We use current knowledge of the natural history of the virus to construct a plausible epidemiological model (Figure 1). We extend a previously published deterministic compartmental model which has provided general insights into the dynamics of the epidemic at a national level for a range of scenarios [13]. The general framework of the model is given in Figure 1 and parameter values are shown in Table 1.

Distributed natural history of infection

The mean latent and infectious periods for SARS-CoV-2 have been estimated as 4.5 days and 3.1 days respectively, using viral load data and the timing of known index and secondary case contacts (Figure 2) [29]. As the probability mass of the latent and infectious period distributions are centred around the mean, we consider that gamma distributions with an integer shape parameter (also known as Erlang distributions), give more realistic waiting times than exponential distributions which have a mode of zero [30, 31, 32].

Transmissibility and infectivity

Estimates of the transmissibility of the virus in the UK at the beginning of the epidemic have ranged from 2.4–3.8 [23, 33, 34], here we assume that $R_0$ at the beginning of the epidemic in the UK population is 2.8. Non-pharmaceutical interventions have been shown to bring the effective reproduction number ($R_t$) below one, and in some settings have led to local elimination of the virus [22, 24].
Testing performed in closed populations suggests that 40-50% of SARS-CoV-2 infections may be asymptomatic [35, 36, 37], while data from contact tracing shows transmission can occur from asymptomatic individuals [38]. We make the assumption that asymptomatic individuals ($I^A$) have 0.5 the infectiousness of symptomatic individuals ($I^S$) [6, 16]. The UK population shows variable contact rates by age [39, 40] and, while studies show mixed results, evidence is accruing that children have a lower susceptibility to acquiring the infection than adults [41, 42, 43]. We assume that children (≤15 years) have 0.4 times the susceptibility of adults [44].

Scenarios for immunity

We allow the duration of immunity to differ for recovered individuals with severe symptoms that are hospitalised ($R^H$) versus those with less severe symptoms that are not hospitalised ($R^N$), as there is evidence from SARS-CoV-2 and other coronaviruses that individuals with milder symptoms may have a lower antibody response (6, 15). The average duration of immunity for hospitalised and non-hospitalised individuals varies by scenario and is described below.

![Flow diagram showing SARS-CoV-2 transmission model outline.](image)

**Figure 1**: Flow diagram showing SARS-CoV-2 transmission model outline. The disease states are susceptible ($S$), exposed ($E$), symptomatic infectious ($I^S$), asymptomatic infectious ($I^A$), hospitalised recovered ($R^H$), and non-hospitalised recovered ($R^N$). Age group specific parameters are indexed by $i$.

### Epidemic transitions for age group $i$ at time $t + 1$ are given by:

$$S_{t+1} = S_t(1 - \lambda_t) + f(R^N_t; o, \omega^N) + f(R^H_t; o, \omega^H)$$  \hspace{1cm} (1)

$$E_{t+1} = E_t + S_t \lambda_t - f(E_t; m, \sigma)$$  \hspace{1cm} (2)

$$I^A_{t+1} = I^A_t + \phi_t f(E_t; m, \sigma) - f(I^A_t; n, \gamma)$$  \hspace{1cm} (3)

$$I^S_{t+1} = I^S_t + (1 - \phi_t) f(E_t; m, \sigma) - f(I^S_t; n, \gamma)$$  \hspace{1cm} (4)

$$R^N_{t+1} = R^N_t + f(I^S_t; n, \gamma) + \left(1 - \frac{p_i}{\phi_t}\right) f(I^A_t; n, \gamma) - f(R^N_t; o, \omega^N)$$  \hspace{1cm} (5)

$$R^H_{t+1} = R^H_t + \frac{p_i}{\phi_t} f(I^S_t; n, \gamma) - f(R^H_t; o, \omega^H)$$  \hspace{1cm} (6)

The function $f(x, \alpha, B)$ represents the Erlang delay distribution within classes $E$, $I^S$, $I^A$, $R^H$ and $R^N$; which is achieved by using $\alpha$ concatenated sub-compartments for each class with rates $B$ between each sub-compartment. If $n$ individuals enter state $X$ at time $t$, by time $t + \tau$ there will be remaining $n(t + \tau) - n(t)$ in $X$, where $g(\tau; \alpha, B)$ gives the cumulative Erlang distribution with (integer) shape parameter $\alpha$ and rate parameter $B$:

$$g(\tau; \alpha, B) = 1 - \sum_{n=0}^{\alpha-1} \frac{1}{n!} e^{-B\tau} (B\tau)^n$$  \hspace{1cm} (7)
The next generation matrix \((K = k_{ij})\) gives the expected number of secondary infections in age group \(i\) resulting from contact with an index case in age group \(j\):

\[
k_{ij} = \frac{\beta}{\gamma} \eta_i c_{i,j} \left( \phi_j v + (1 - \phi_j) \right)
\]

The basic reproduction number \((R_0)\) is given by the spectral radius \(\rho(K)\) which is the largest absolute eigenvalue of \(K\). The force of infection acting on age group \(i\) at time \(t + 1\) is given by:

\[
\lambda_{t+1} = \beta \eta_i \sum_{j=1}^{N_a-1} c_{i,j} \frac{I_{j,t}^S + I_{j,t}^A}{N_j} \nu
\]

where \(c_{i,j}\) is the average number of daily contacts in the population between age groups \(j\) and \(i\); \(N_a\) is the number of discrete age groups \((N_a = 15)\); and \(N_j\) gives the population size of age group \(j\). As we specify the value of \(R_0\), the transmission parameter \(\beta\) is left as a free parameter which is scaled to the correct value.

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Symbol</th>
<th>Estimate(s)</th>
<th>Details</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic reproduction number</td>
<td>(R_0)</td>
<td>2.8</td>
<td>-</td>
<td>Key assumption</td>
</tr>
<tr>
<td>Latency period mean</td>
<td>(\sigma)</td>
<td>4.5 days</td>
<td>-</td>
<td>([29, 46, 47])</td>
</tr>
<tr>
<td>Latency period shape</td>
<td>(m)</td>
<td>4</td>
<td>-</td>
<td>([29, 46, 47])</td>
</tr>
<tr>
<td>Infectious period mean</td>
<td>(\gamma)</td>
<td>3.1 days</td>
<td>-</td>
<td>([29, 46, 47])</td>
</tr>
<tr>
<td>Infectious period shape</td>
<td>(n)</td>
<td>2</td>
<td>-</td>
<td>([29, 46, 47])</td>
</tr>
<tr>
<td>Immune duration mean non-hospitalised</td>
<td>(\omega^N)</td>
<td>(\infty, 365, 180, 90) days</td>
<td>Varies by scenario</td>
<td>([11, 45, 48])</td>
</tr>
<tr>
<td>Immune duration mean hospitalised</td>
<td>(\omega^H)</td>
<td>(\infty, 365) days</td>
<td>Varies by scenario</td>
<td>([11, 45, 48])</td>
</tr>
<tr>
<td>Immune duration shape</td>
<td>(o)</td>
<td>2</td>
<td>Centres distribution around mean</td>
<td>([11])</td>
</tr>
<tr>
<td>(P(\text{asymptomatic} \mid \text{infection}))</td>
<td>(\phi_i)</td>
<td>(\leq 15) yrs 0.75, (&gt;15) yrs 0.5</td>
<td>Varies by age group (i)</td>
<td>([41])</td>
</tr>
<tr>
<td>(P(\text{hospitalisation} \mid \text{infection}))</td>
<td>(p_i)</td>
<td>0–0.26</td>
<td>Varies by age group (i)</td>
<td>([16, 36])</td>
</tr>
<tr>
<td>Effective reproduction number</td>
<td>(R_t)</td>
<td>1.1, 0.8</td>
<td>During lockdown</td>
<td>([49])</td>
</tr>
<tr>
<td>Contact matrix</td>
<td>(C \text{ or } c_{ij})</td>
<td>Varies by age group</td>
<td>BBC survey</td>
<td>([39, 40])</td>
</tr>
<tr>
<td>Relative infectiousness of asymptomatic cases</td>
<td>(\upsilon)</td>
<td>0.5</td>
<td>-</td>
<td>([40])</td>
</tr>
<tr>
<td>Relative age susceptibility</td>
<td>(\eta_i)</td>
<td>(\leq 15) yrs 0.4, (&gt;15) yrs 1</td>
<td>-</td>
<td>([44])</td>
</tr>
</tbody>
</table>

Table 1: Summary of parameter values used in the modelled scenarios of SARS-CoV-2 transmission in the United Kingdom.

**Immunological scenarios**

Using data and timing of events from the UK epidemic, we explore four scenarios with varying average durations of immunity to SARS-CoV-2 (Figure 2).

- **S1. Permanent:** Where immunity is lifelong for both hospitalised \((R^H)\) and non-hospitalised \((R^N)\) cases.
- **S2. Waning (12 months):** Where immunity is lifelong for hospitalised cases and has an average duration of 365 days for non-hospitalised cases.
- **S3. Waning (6 months):** Where immunity is lifelong for hospitalised cases and has an average duration of 180 days for non-hospitalised cases.
- **S4. Short-lived:** Where immunity lasts, on average, 365 days for hospitalised cases and 90 days for non-hospitalised cases.
UK-specific parameterisation

All scenarios are initialised with 200 infected individuals in early February 2020. Intervention measures are initiated on 23rd March (date the UK nationwide lockdown started), with an immediate reduction in the effective reproduction number (expected number of secondary cases from an index case at time t; $R_t$) to 1.1 for a three week period, followed by a further reduction in $R_t$ to 0.8 until lockdown measures are eased on 15th June [49]. After this time, $R_t$ is brought to 0.9, 1, 1.1 or 1.2 until February 2022. We considered the majority of our analysis over a, relatively short, two year period to explore the epidemic up to a secondary peak; beyond this point the dynamics are likely to be altered depending on further interventions or changes to $R_t$. As we simulate disease dynamics over a relatively short period of time, we do not consider demography (births and deaths) or transitions between age classes (ageing). To obtain equilibrium values, we simulated epidemic trajectories for up to five years.

The UK contact matrix (average daily contacts between an individual in age group $j$ with individuals in age group $i$) comes from a ‘citizen science’ project for the BBC, in which individuals in the UK population provided detailed information on their daily contacts in the home, in the workplace, at school and in other settings [39, 40]. The contact matrix is altered to account for changes to contact patterns during and after the main intervention period [47]. During the lockdown, home; work; school; and other contacts are reduced to 0.8, 0.3, 0.1 and 0.2 respectively of their baseline values. This reflects the school closures for all children, except for those of key workers, and that workers were encouraged to work from home. Reduction in home contacts accounts for the absence of visitors to the home during the lockdown. In the post-lockdown phase, home; work; school; and other contacts are scaled to 1, 0.8, 0.85 and 0.75, respectively, of their baseline values to reflect limited social distancing measures that are likely to be in place until at least the end of 2021.

Analysis was performed in R version 3.6.3. We present figures from model output in the text to the nearest thousand. Code is available at https://github.com/tc13/covid-19-immunity.

![Figure 2: Probabilities for time spent in each state given gamma distributed rates of removal. A. Proportion of individuals in exposed and infectious classes since time from infection. Time exposed and time infectious have mean durations of 3.1 and 4.5 days respectively. B. Proportion of individuals immune since recovery, where time immune has mean durations of 90, 180 or 365 days depending on the scenario.](https://github.com/tc13/covid-19-immunity)

Results

Age structure

The epidemic is driven by the rate of infectious contacts between individuals in different age groups. This is described by the next generation matrix in which the average number of secondary cases generated by an index case in age group $j$ is the summation of row $j$ (Equation 8 & Figure 3). At the beginning of the epidemic, when SARS-CoV-2 is spreading rapidly, all age groups are involved in transmission; in particular those aged 20–39 years. An index case in the 20–24 age group, for instance, is expected to generate an average of 3.1 secondary cases at baseline. As lockdown measures come into force this dramatically reduces the expected number of secondary cases due to fewer contacts and a lower probability of infection given contact. The average number of secondary cases from an individual aged 20–24 during lockdown drops to 0.9 and the transmission parameter...
Figure 3: Next generation matrix ($K = k_{ij}$) showing the number of secondary cases generated by an index case from age group $j$ (rows) in age group $i$ (columns). The matrices are shown for different time points; at baseline before the implementation of interventions; during the lockdown period; and in the post-lockdown period when the effective reproduction number ($R_t$) rises from 0.9–1.2. The average number of secondary cases generated by an index case from age group $j$ is the summation of row $j$.

$\beta$, which captures the probability of infection given contact, is decreased from 0.13 at baseline to 0.11. In the post-lockdown period daily contacts are increased to a higher proportion of their baseline values (see Methods); in order to keep the reproduction number equal to the dominant eigenvalue of the next generation matrix, $\beta$ is consequently reduced to 0.05 when $R_t = 0.9$ and to 0.07 when $R_t = 1.2$. This implies that, to maintain $R_t$ below one when more contacts are occurring in the population post-lockdown, the probability that contact results in infection will need to be reduced.

Infection dynamics

For the first 130 days until the end of the lockdown, the infection dynamics are equivalent across the four immunity scenarios S1–S4 (Figure 4, panels A, C, E & G). After this time the dynamics depend on both the rate at which recovered individuals lose immunity and become re-susceptible, and the post-lockdown $R_t$.

Given our model and parameters, on the first day the intervention is imposed (23rd March 2020) there are 96,000 new SARS-CoV-2 cases, which is within the 95% credible interval (CrI) of new cases estimated for the UK on that day (95% CrI 54,000–155,000 [50]), and 124,000 people are infectious (infected compartments $I_A + I_S$) on this date. From 16th February until 23rd March there are 717,000 cumulative cases across all age groups and 680,000 in adults ≥19 years, which narrowly exceeds the credible interval for an estimate of cumulative cases in this period (95% CrI 266,000–628,000 [50]). When most of the lockdown measures were eased in June, 5.5% of the total population and 6.8% of adults aged ≥19 years have immunity to SARS-CoV-2 (in recovered classes $R_H$ and $R_N$), which is comparable to estimates of antibody levels in the UK population, estimated as 6.8% of
blood donors on 24th May 2020 (95% confidence interval 5.2–8.6%; individuals ≥18 years [25]).

Secondary peak in infections

A secondary peak in infections is expected in spring 2021 where $R_t = 1.1$ or winter 2020 where $R_t = 1.2$ (Figure 4, panels E & G). The height of the secondary peak is determined by the rate at which immunity is lost. In our worst case scenario (S4: short-lived immunity) where immunity lasts an average of three months for non-hospitalised patients, a year for hospitalised patients and $R_t$ following lockdown is 1.2, then the secondary peak will exceed the initial peak with a maximum of 409,000 infectious individuals and 133,000 daily new cases in December 2020. This is nearly triple the number of new cases compared with scenario S1 where immunity is permanent; the maximum number of infectious individuals in the secondary peak is 137,000 and there are 45,000 daily new cases (Figure 4, panels A, C, E & G). We note that the timing of the secondary peak in infection curves across immunological scenarios are closely synchronised and in autumn 2020. This synchrony and timing is also observed during the epidemic when values of $R_t$ post-lockdown are greater than 1.2 (explored for values of $R_t$ from 1.3 to 2.0).

When $R_t$ following lockdown is 1.1, the differences between the scenarios is even more pronounced with a six-fold difference in the height of the secondary peak of infectious individuals between a scenario of permanent immunity and one of short-lived immunity. When immunity wanes rapidly, a secondary peak is observed in April 2021 with a maximum of 161,000 infectious individuals and 52,000 daily new cases. By contrast when immunity is permanent, the number of new infections slowly decays rather than accelerates, and there are projected to be only 24,000 infectious individuals and 5,000 daily new cases in April 2021 (Figure 4E).

Population immunity

Dynamics of population immunity (recovered compartments $R^H + R^N$) are similarly shaped by the expected duration of antibodies against SARS-CoV-2 and the post-lockdown $R_t$.

Immunity decays from midway through the lockdown period in scenarios S2–S4 of waning (12 or 6 months) and short-lived immunity and resurges following a secondary wave of infection if $R_t > 1$ (Figure 4, panels F & H). After lockdown, a fall in the proportion of the population immune to the virus is observed until autumn 2020 for all values of $R_t$, after which point the secondary peak, if $R_t > 1$, causes the proportion of the population immune to rise again. This suggests that longitudinal serological surveys to detect waning immunity would be most informative when conducted in the period June–September 2020.

Consequences of age structure

The large differences in the heights of the secondary peaks when $R_t > 1$ between immunological scenarios (Figure 4, panels A, C, E & G) can be explained by the heterogeneity in transmission (see the next generation matrix in Figure 5) for scenarios S1 & S4 of permanent and short-lived immunity where $R_t = 1.2$ following lockdown. A higher proportion of individuals aged between 20–39 are infected early in the epidemic, and this leads to 10.5–12.6% for scenarios S1 & S4 of permanent and short-lived immunity and resurges following a secondary wave of infection if $R_t > 1$ (Figure 4, panels F & H). After lockdown, a fall in the proportion of the population immune to the virus is observed until autumn 2020 for all values of $R_t$, after which point the secondary peak, if $R_t > 1$, causes the proportion of the population immune to rise again. This suggests that longitudinal serological surveys to detect waning immunity would be most informative when conducted in the period June–September 2020.

Longer term dynamics: extinction or endemic equilibrium

We explored the impact of waning immunity and $R_t$ on the equilibrium values for the different simulations over a longer, five year, period until February 2025 (Table 2). If the post-lockdown $R_t$ is suppressed below one following lockdown, then the differences in immunity will have less impact on the longer-term infection dynamics, assuming no imported cases, as transmission of SARS-CoV-2 becomes unsustainable and the virus reaches extinction between April–November 2021 depending on the immunity scenario. In simulations where $R_t$ equals one, if immunity is permanent then the epidemic becomes extinct in May 2022. When immunity wanes there is no secondary peak (Figure 4C), however the infections persist at a low level of endemicity equivalent to 106, 233 and 1,168 daily cases in immunity scenarios S2–S4, respectively. For larger values of $R_t$, and where immunity wanes, the system oscillates with subsequent peaks of infection over the next five years until a steady state is reached. We find that, if $R_t = 1.2$ post-lockdown and immunity is short-lived, there is the potential for over 76,000 new cases daily; 6,000 hospitalisations; and 1,000 intensive care unit (ICU) admissions (calculated...
as 17% of all hospitalised cases [51] at endemic equilibrium (January 2025), which would be sufficient to overwhelm contact tracing services and ICU capacity [52, 53].

Table 2: Values at equilibrium from the modelled scenarios for SARS-CoV-2 in the United Kingdom, explored over a five year horizon (February 2020 to February 2025). ¹Effective reproduction number of SARS-CoV-2 following after lockdown. ²Assumed duration of immunity for hospitalised and non-hospitalised individuals, see Methods for details of scenarios S1–S4. ³Number of individuals newly infected with SARS-CoV-2 that enter the exposed E state. ⁴Number of symptomatic individuals with SARS-CoV-2 that enter the recovered hospitalised R₅ state. ⁵Number of hospitalised individuals admitted to intensive care units, under the assumption that 17% of hospitalised cases in the UK require care in high dependency units [51]. ⁶Either when the number of daily new cases drops below one (extinction), or when the daily new cases are the same integer value over a sustained period (endemic equilibrium). If models take longer than five years to reach a steady state, the values are reported for the last day on 31st January 2025.

Discussion

Despite only 6% of the adult UK population having immunity against SARS-CoV-2 in our simulation at the end of the lockdown, the modelled scenarios suggest that, if this acquired immunity wanes over time, there are substantive differences to the subsequent infection dynamics. Waning immunity impacts on the height of the secondary peak and, in the absence of future interventions, establishes the virus at levels of endemic equilibrium that could overwhelm contact tracing services and ICU capacity [52, 53].

We predict that surveys to detect waning immunity at the population level would be most effective when carried out in the period between the end of lockdown and autumn 2020, as after this point an upsurge in cases is expected that will increase the proportion of the population with antibodies to SARS-CoV-2. In particular, this will allow evaluation of whether specific antibodies generated against the virus are short-lived if reductions in antibody prevalence are observed at the population level.

We find that transmission is driven disproportionately by individuals of working age, and subsequently a higher proportion of individuals aged 20–39 years become infected early in the pandemic and subsequently develop antibodies (Figures 3 & 5). This prediction is borne out by serological data from Switzerland, which showed that individuals aged 20–49 years were significantly more likely to be seropositive in May 2020 compared with younger and older age groups [54]. We postulate that ‘key workers’ in the UK population who have continued to work during the lockdown are more likely to have antibodies against SARS-CoV-2. Higher immunity among individuals of working age has the effect of slowing the subsequent epidemic when immunity is permanent. Conversely, when immunity wanes, previously infected individuals of working age re-join the susceptible pool and so contribute again to transmission; leading to a high growth rate and a larger secondary peak of infected cases. In these circumstances, efforts to suppress transmission will be challenging in the absence of a transmission-blocking vaccine [15]. We note that the model structure developed here is capable of simulating the impact of vaccination with a vaccine that provides temporary transmission-blocking immunity, and could be used to predict the optimal timing for booster shots.
Figure 4: Projections from immunity scenarios S1–4 with post-lockdown $R_t$ ranging from 0.9–1.2. Left panels show the number of infected, both asymptomatic and symptomatic ($I^A + I^S$), with SARS-CoV-2 in the UK population over time. Right panels show the proportion of the UK population with immunity (compartments $R^H + R^N$). Dashed vertical lines indicate the lockdown period; 23rd March–15th June 2020.
The projected trajectory of the epidemic after lockdown is highly sensitive to the effective reproduction number, with model behaviour for values of $R_t$ slightly above or below one displaying qualitatively different dynamics (Figure 4). This shows the importance of timely and accurate estimates of $R_t$ to inform control strategies, and ensuring widespread community testing and contact tracing is in place. Our calculations show that to suppress $R_t$ below one when contact rates rise to a higher fraction of baseline (pre-lockdown) values, the probability of infection given contact (represented here by the $\beta$ parameter), must drop by around half. Interventions that have the potential to reduce the probability of infection include social distancing; regular hand washing; and the wearing of face masks outside the home [55].

Our study reinforces the importance of better understanding SARS-CoV-2 immunity among recovered individuals of different ages and disease severity. In scenarios where immunity wanes and $R_t$ following lockdown is greater than one, the SARS-CoV-2 epidemic never reaches extinction due to herd immunity, but rather the number of infected cases oscillates with subsequent waves of infection before reaching endemic equilibrium (Table 3). Even in simulations where the reproduction number only narrowly exceeds one, if immunity wanes over an average of one year for severe cases and three months for non-severe cases, this is projected to lead to an equilibrium state of over 40,000 daily new cases and 200 daily admissions to intensive care. Policy strategies aiming to achieve herd immunity are therefore risky [5], as if SARS-CoV-2 antibodies do wane over time, then a herd immunity threshold can never be reached in the absence of a vaccine that provides permanent immunity [21]. The establishment of an endemic equilibrium state is dependent on no future interventions or changes to $R_t$, which we consider unlikely as policy makers and public health agencies are likely to react to future outbreaks with localised control measures.

One of the strengths of our study is that the model is calibrated to key features of the UK epidemic. While we did not explicitly fit to data, new cases at the start of the lockdown; cumulative cases between February
and March; and the proportion of the adult population with antibodies to SARS-CoV-2 are highly comparable between our output and current estimates [29, 50]. We used contact matrices from a comprehensive study of contact patterns in the UK population [59] in addition to demographic data from the Office for National Statistics, to give our simulations the best chance of capturing realistic age-specific transmission patterns in the UK population.

Plausible estimates on which to base expectations for the duration of immunity are sparse in the current literature. Rosado et al. estimated that antibodies could wane in 50% of recovered individuals after one year [48], which is similar to the estimated duration of immunity against seasonal circulating coronaviruses [13]. Even with this consideration, there are many probability distributions that can be used to capture a median duration of immunity, and our selection of an Erlang distribution with a shape parameter of two is somewhat arbitrary. Our assumptions on the duration of the latent and infectious periods are more closely informed by estimates from data [29, 46, 47]. We made the decision to capture the expected duration of these states as Erlang distributions rather than the, more conventional, exponential distribution. This has the benefit of closely replicating fitted gamma or log-normal distributions within a compartmental model [50], and has important implications for the dynamics of the epidemic [58, 57]. We make a number of assumptions regarding the natural history of the virus, such as the relative susceptibility of children compared with adults and the relative infectiousness of symptomatic versus asymptomatic cases based on the current literature [44, 11]. Future empirical studies are likely to add to and further refine these epidemiological parameters. After we completed the analysis, a study of 37 asymptomatic individuals in China were found to have a longer period of viral shedding when compared with symptomatic individuals [10]. While viral shedding is not necessarily indicative of transmission potential [7], if these findings are replicated in larger studies this may suggest a need to use different durations of infectiousness for asymptomatic and symptomatic infections in subsequent models.

We have aimed to capture future infection dynamics at a national level in the UK under a range of scenarios. Our analysis is limited by not considering regional differences in transmission rates, for instance through a patch (metapopulation) model [31], or a stochastic approach that allows for local extinction events [21]. There are no deaths in our model, either from demography or infection. Accounting for mortality would mainly affect dynamics in the oldest age group (over 70 years) [16, 51], as the higher probability of disease-induced mortality would prevent a substantial build up of immunity (Figure 5D). We also do not explicitly consider transmission in settings such as hospitals or care homes, although such dynamics may be captured indirectly through the contact matrix. Given the simplicity of the model structure, we advise against treating the output as an exact prediction of the future. In addition to the limitations listed above, the epidemic trajectory will be substantially altered by any future interventions such as a return to full lockdown conditions, or intensive contact tracing and isolation [13, 58].

Acknowledgements

This work was undertaken as a contribution to the Rapid Assistance in Modelling the Pandemic (RAMP) initiative, coordinated by the Royal Society. The authors thank Prof. Julia Gog and Dr. Rosalind Eggo for their helpful comments and input.

Disclosure statement

The authors declare no conflicts of interest. No other authors have been paid to write this article.

Open access

Anyone can share and adapt this article provided they give credit and link to this article’s DOI and CC BY 4.0 licence (creativecommons.org/licenses/by/4.0).

Funding

TC is funded by a Sir Henry Wellcome Postdoctoral Fellowship from the Wellcome Trust (reference 215919/Z/19/Z). ELD, TCDL, DA, AB, TMP and TDH gratefully acknowledge funding of the NTD Modelling Consortium by the Bill & Melinda Gates Foundation (BMGF) (grant number OPP1184344). Views, opinions, assumptions or any other information set out in this article should not be attributed to BMGF or any person connected with them. TMP’s PhD was supported by the Engineering & Physical Sciences Research Council, Medical Research Council and University of Warwick (grant number EP/LO15374/1). All funders had no role in the study design, collection, analysis, interpretation of data, writing of the report, or decision to submit the manuscript for publication.
References


