Supplementary Materials for

A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2

Tom Britton*, Frank Ball, Pieter Trapman

*Corresponding author; E-mail: tom.britton@math.su.se

Published 23 June 2020 on *Science* First Release
DOI: 10.1126/science.abc6810

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/cgi/content/full/science.abc6810/DC1)

Materials and Methods

A deterministic SEIR model and the fraction of the population infected

In this supplementary information we describe the deterministic SEIR (Susceptible, Exposed, Infectious, Removed) epidemic model in a population partitioned by age and activity level. For reasons of notational convenience we label the types (the combination of age and activity level) from 1 to m, where m is the product of the number of age classes and the number of activity levels. A more detailed exposition than the one presented here can be found in [(*15*), Sections 5.5 and 6.2].

We assume that for all $j \in \{1, \cdots, m\}$ the population consists of n_j people of type j. We set $n = \sum_{j=1}^{m} n_j$ and $\pi_j = n_j/n$. We assume that the population is large and closed, in the sense that we do not consider births, deaths (other than possibly the deaths caused by the infectious disease) and migration. Throughout the epidemic, n_i is fixed, so people who die from the infectious disease are still considered part of the population. For $j, k \in \{1, \cdots, m\}$, every given person of type j makes infectious contacts with every given person of type k independently at rate $\alpha a_{jk}/n$. If at the time of such a contact the type-j person is infectious and the type-k person is susceptible then the latter becomes latently infected (Exposed). People of the same type may infect each other, so a_{jj} may be strictly positive. Because the definition of an infectious contact includes that the contact leads to transmission of the disease, it is not necessarily the case that a_{jk} is equal to a_{kj}. The parameter α is a scaling parameter, used to quantify the impact of control measures in the main paper, without measures α is set equal to 1. Exposed individuals become Infectious at constant rate σ and infectious individuals recover or die (are Removed) at constant rate μ. The rates of becoming infectious and removal are assumed to be independent of type. It is straightforward to extend the model to make those rates age and/or activity level dependent.

In the described multi-type SEIR model, the expected number of people of type k that are infected by an infected person of type j during the early stages of the epidemic is $n_k \times (\alpha a_{jk}/n) \times (1/\mu) = \pi_k \alpha a_{jk}/\mu$, where $1/\mu$ is the expected duration of an infectious period. The next-generation matrix M has (for $j, k \in \{1, \cdots, m\}$) as element in the j-th row and m-th column the quantity $\pi_k \alpha a_{jk}/\mu$. The basic reproduction number R_0 is defined as the largest eigenvalue of M; it is necessarily real and positive. If $R_0 > 1$, then a large outbreak is possible with positive probability, while if $R_0 \leq 1$ an outbreak stays small with probability 1.

We set $S_j(t)$ to be the number of people of type j that are susceptible to the disease at time t, $E_j(t)$ the number of people of type j that are latently infected, $I_j(t)$ the number of infectious people of type j and $R_j(t)$ the number of removed people of type j ($j \in \{1, \cdots, m\}$). Note that $S_j(t) + E_j(t) + I_j(t) + R_j(t) = n_j = \pi_j n$ for all $t \geq 0$, because the population is closed. Again for $j \in \{1, \cdots, m\}$, we define $s_j(t) = S_j(t)/n_j$,

$e_j(t) = E_j(t)/n_j$, $i_j(t) = I_j(t)/n_j$ and $r_j(t) = R_j(t)/n_j$.

Theory on Markov processes [(*16*), Chapter 11; see also (*15*), Section 5.5, for the single type counterpart] gives that for large n the above model can be described well by the following system of differential equations (again for $j \in \{1, \cdots, m\}$):

$$\begin{aligned}
\dot{s}_j(t) &= -\tfrac{1}{n_j}\sum_{k=1}^{m} \alpha \frac{a_{kj}}{n} S_j(t) I_k(t) &&= -\sum_{k=1}^{m} \alpha \pi_k a_{kj} s_j(t) i_k(t), \\
\dot{e}_j(t) &= \tfrac{1}{n_j}\left(\sum_{k=1}^{m} \alpha \frac{a_{kj}}{n} S_j(t) I_k(t) - \sigma E_j(t)\right) &&= \sum_{k=1}^{m} \alpha \pi_k a_{kj} s_j(t) i_k(t) - \sigma e_j(t), \\
\dot{i}_j(t) &= \tfrac{1}{n_j}\left(\sigma E_j(t) - \mu I_j(t)\right) &&= \sigma e_j(t) - \mu i_j(t), \\
\dot{r}_j(t) &= \tfrac{1}{n_j}\mu I_j(t) &&= \mu i_j(t).
\end{aligned}$$

To be complete, in the main text, we use when analysing the time-dependent behaviour of an epidemic that for all $j \in \{1, \cdots, m\}$, $s_j(0) = 1 - \epsilon$, $e_j(0) = \epsilon$ and $i_j(0) = r_j(0) = 0$. In the analysis below we do not impose specific assumptions on the initial conditions.

The epidemic will ultimately go extinct, because the population is closed, so for all $j \in \{1, \cdots, m\}$ we have that $e_j(t) \to 0$ and $i_j(t) \to 0$ as $t \to \infty$. Thus $s_j(t) + r_j(t) \to 1$ as $t \to \infty$. Furthermore $s_j(t)$ is non-increasing, so $s_j(\infty) = \lim_{t\to\infty} s_j(t)$ exists.

It can be shown in the spirit of [(*15*), Equation (6.2)] that for $j \in \{1, \cdots, m\}$,

$$\frac{s_j(\infty)}{s_j(0)} = \exp\left[-\lambda \sum_{k=1}^{m} a_{kj}\pi_k \left(1 - r_k(0) - s_k(\infty)\right)/\mu\right]. \tag{1}$$

To understand this identity we observe first that $\frac{s_j(\infty)}{s_j(0)}$ is the fraction of initially susceptible people of type j who escape the epidemic, while the sum in the right-hand side can be written as

$$\sum_{k=1}^{m} n\pi_k \left(1 - r_k(0) - s_k(\infty)\right) \times \lambda a_{kj}/n \times \frac{1}{\mu} = \sum_{k=1}^{m} (n_k - R_k(0) - S_k(\infty)) \times \lambda a_{kj}/n \times \frac{1}{\mu}.$$

In words the summands read as the number of people of type k that were infectious at some moment during the epidemic, times the rate at which a type-k person makes infectious contacts with someone of type j, times the expected time an infected person is infectious. In other words, the right hand side is the cumulative force of infection during the entire epidemic acting on a person of type j. Standard theory on epidemics gives that minus the natural logarithm of the probability that a given initially susceptible person of type j avoids infection is the cumulative force of infection acting on that person. Thus (1) gives that the fraction of initially susceptible people that are ultimately still susceptible is equal to the probability that a given initially susceptible person avoids infection.

If the epidemic is initiated by few infectives in a large population then, conditional upon a large outbreak occurring, the final fractions of initially susceptible people of the

different types satisfy (1) with $s_j(0) = 1$ and $r_j(0) = 0$ for all $j \in \{1, \cdots, m\}$. In the main text, this special case is used to calculate h_D in Table 1 and the final fractions infected in Table 2.

The population matrix

In the main text we analyse an age structured population. Contact intensities between different age groups we took from (*14*). The age groups are 0-5, 6-12, 13-19, 20-39, 40-59 and 60+. The contact matrix, i.e. the matrix with elements $\{a_{jk}; j, k \in \{1, \cdots, 6\}\}$ is taken from Table 1 of (*14*). Note that the expected number of contacts from a person of type j with people of type k is $n_k a_{jk}/n = \pi_k a_{jk}$. Therefore we divide the elements of Table 1 by the corresponding π_k to obtain the contact matrix. (The values of π_j , $j \in \{1, \cdots, m\}$, are obtained using Appendix Table 1 of (*14*); $\pi_1 = 0.0725$, $\pi_2 = 0.0866$, $\pi_3 = 0.1124$, $\pi_4 = 0.3323$, $\pi_5 = 0.2267$, $\pi_6 = 0.1695$.) We further multiply this matrix by a constant such that the largest eigenvalue is equal to 2.5, the value we have chosen for R_0. The contact matrix is

$$\begin{pmatrix} 2.2257 & 0.4136 & 0.2342 & 0.4539 & 0.2085 & 0.1506 \\ 0.4139 & 3.6140 & 0.4251 & 0.4587 & 0.2712 & 0.1514 \\ 0.2342 & 0.4257 & 2.9514 & 0.6682 & 0.4936 & 0.1972 \\ 0.4539 & 0.4592 & 0.6676 & 0.9958 & 0.6510 & 0.3300 \\ 0.2088 & 0.2706 & 0.4942 & 0.6508 & 0.8066 & 0.4341 \\ 0.1507 & 0.1520 & 0.1968 & 0.3303 & 0.4344 & 0.7136 \end{pmatrix}.$$

As explained in the main text we can use this matrix to generate the 18 by 18 contact matrix for the model in which we take both age and activity level into account.

Gradual lifting of restrictions

In the main text we studied effects of lifting restrictions of different levels α on June 30 (day 135) going back to the situation of no restrictions corresponding to setting α back to 1. Below are the corresponding plots but where restrictions are relaxed gradually (linearly) between June 1 (day 105) and August 31 (day 195). In Figure S1 we plot the fraction of infectious individuals.

The plot looks quite similar to that of Figure 1 in the main text except that the purple curve with highest restrictions no longer has a pronounced second wave. The reason for this is that now restrictions are lifted slowly and gradually during a 3 month period rather than directly back to normal at a single time point.

In Figure S2 the corresponding plot for cumulative fraction of infected over time is given. Compared to Figure 2 the main difference is that the purple curve (severe restrictions) has fewer finally infected since there is no longer such a large overshoot above $h_D = 43\%$ caused by a second wave.

Finally, in Figure S3 we consider the situation when preventive measures with level α are implemented 30 days after introduction of the disease and relaxed (so α returns to 1) at time $t > 30$. The parameters are again chosen so that $R_0 = 2.5$ when $\alpha = 1$. The graphs show the effective R_0 (incorporating disease-induced immunity) as a function of time t, for four different choices of preventive level α. Thus all four curves coincide until day 30. The effective R_0 with no preventive measures ($\alpha = 1$) reaches the critical value of one on about day 57 (mid-April), whilst that for $\alpha = 0.8$ does so on about day 68 (April 23). The stronger preventive measures ($\alpha = 0.6$ and $\alpha = 8/15$) are such that herd immunity is never reached even if they are retained indefinitely.

Fig. S1.

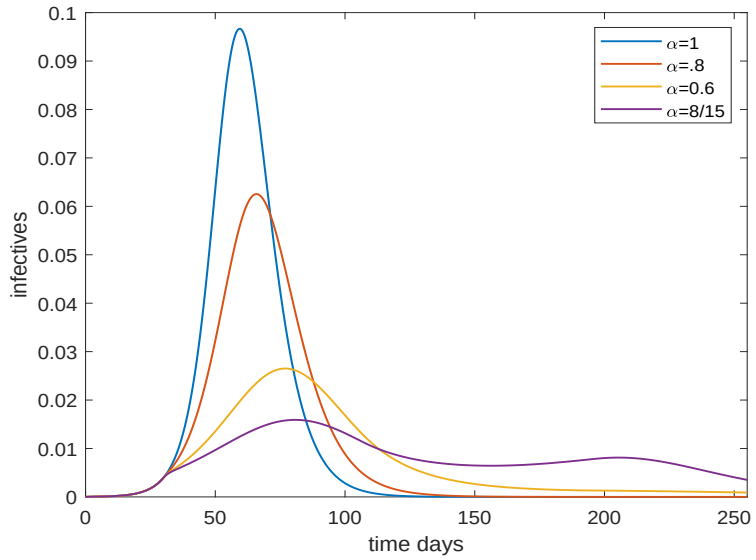

Figure S1: Plot of the overall fraction infected over time for the age and activity structured community with $R_0 = 2.5$, for four different preventive levels inserted March 15 (day 30) and lifted gradually between June 1 (day 105) and August 31 (day 195). The blue, red, yellow and purple curves corresponds to no, light, moderate and severe preventive measures, respectively.

Fig. S2.

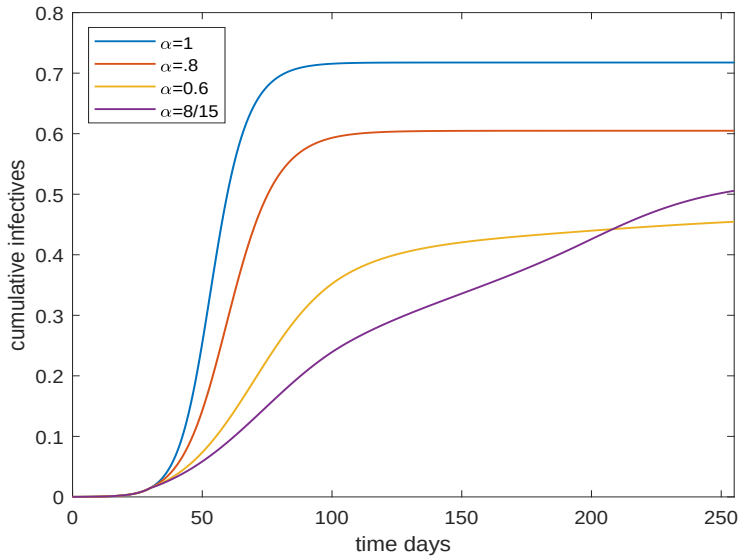

Figure S2: Plot of the overall fraction infected over time for the age and activity structured community with $R_0 = 2.5$, for four different preventive levels inserted March 15 (day 30) and lifted gradually between June 1 (day 105) and August 31 (day 195). The blue, red, yellow and purple curves corresponds to no, light, moderate and severe preventive measures, respectively.

Fig. S3.

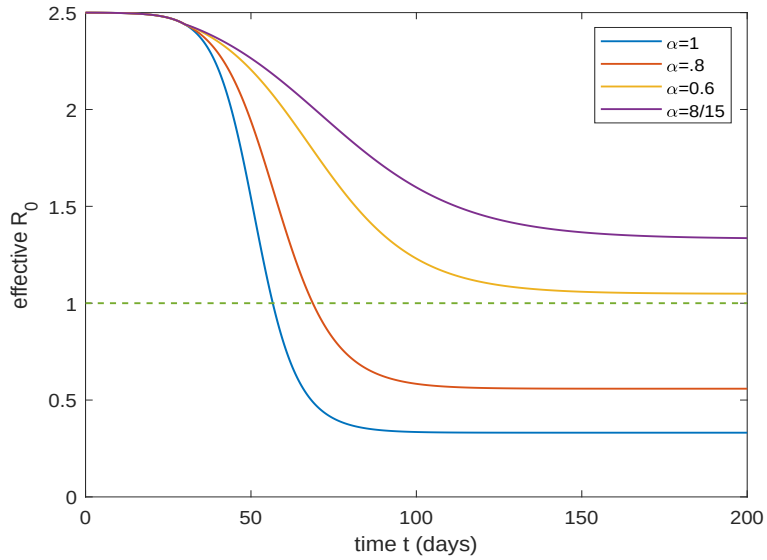

Figure S3: Plot of the effective reproduction number (incorporating disease-induced immunty) if restrictions for different α are put in place Day 30 and relaxed on day $t > 30$.

Table S1

Table S1: Final outcome fractions infected in different groups for the model with only one activity structure assuming preventive measures put in place such that $\alpha = \alpha_*$ just barely reaching herd immunity, for $R_0 = 2.0, 2.5$ and 3.0. Fractions infected are given as percentages.

R_0	Low activity	Average activity	High activity
2.0	19.9	35.9	58.9
2.5	25.8	44.9	69.7
3.0	30.4	51.5	76.5

References and Notes

1. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, H. Coupland, T. A. Mellan, H. Zhu, T. Berah, J. W. Eaton, P. N. P. Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, M. Baguelin, I. Blake, A. Boonyasiri, O. Boyd, L. Cattarino, C. Ciavarella, L. Cooper, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, S. van Elsland, R. FitzJohn, H. Fu, K. Gaythorpe, L. Geidelberg, N. Grassly, W. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati-Gilani, P. Nouvellet, K. Parag, I. Siveroni, H. Thompson, R. Verity, E. Volz, C. Walters, H. Wang, Y. Wang, O. Watson, P. Winskill, X. Xi, C. Whittaker, P. G. T. Walker, A. Ghani, C. A. Donnelly, S. Riley, L. C. Okell, M. A. C. Vollmer, N. M. Ferguson, S. Bhatt, Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries (Imperial College London, 2020); https://doi.org/10.25561/77731.

2. Public Health Institute of Sweden, Estimates of the peak-day and the number of infected individuals during the covid-19 outbreak in the Stockholm region, Sweden February–April 2020 [in Swedish] (2020); www.folkhalsomyndigheten.se/contentassets/2da059f90b90458d8454a04955d1697f/skattning-peakdag-antal-infekterade-covid-19-utbrottet-stockholms-lan-februari-april-2020.pdf.

3. Instituto de Salud Carlos II, Second round national sero-epidemiology study of SARS-COV-2 infection in Spain [in Spanish] (2020); https://www.mscbs.gob.es/ciudadanos/ene-covid/docs/ESTUDIO_ENE-COVID19_SEGUNDA_RONDA_INFORME_PRELIMINAR.pdf.

4. N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. T. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, A. C. Ghani, Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand (Imperial College London, 2020); https://doi.org/10.25561/77482.

5. W. Bock, B. Adamik, M. Bawiec, V. Bezborodov, M. Bodych, J. P. Burgard, T. Goetz, T. Krueger, A. Migalska, B. Pabjan, T. Ozanski, E. Rafajlowicz, W. Rafajlowicz, E. Skubalska-Rafajlowicz, S. Ryfczynska, E. Szczurek, P. Szymanski, Mitigation and herd immunity strategy for COVID-19 is likely to fail. medRxiv 2020.03.25.20043109 [Preprint]. 5 May 2020; https://doi.org/10.1101/2020.03.25.20043109.

6. H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, S. Cauchemez, Estimating the burden of SARS-CoV-2 in France. *Science* 10.1126/science.abc3517 (2020). doi:10.1126/science.abc3517

7. O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics* (Princeton Univ. Press, 2013).

8. J. Wallinga, P. Teunis, M. Kretzschmar, Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* **164**, 936–944 (2006). doi:10.1093/aje/kwj317 Medline

9. R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001). doi:10.1103/PhysRevLett.86.3200 Medline

10. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, 1991).

11. M. J. Ferrari, S. Bansal, L. A. Meyers, O. N. Bjørnstad, Network frailty and the geometry of herd immunity. *Proc. Biol. Sci.* **273**, 2743–2748 (2006). doi:10.1098/rspb.2006.3636 Medline

12. F. Ball, T. Britton, O. Lyne, Stochastic multitype epidemics in a community of households: Estimation and form of optimal vaccination schemes. *Math. Biosci.* **191**, 19–40 (2004). doi:10.1016/j.mbs.2004.05.001 Medline

13. M. G. M. Gomes, R. M. Corder, J. G. King, K. E. Langwig, C. Souto-Maior, J. Carneiro, G. Goncalves, C. Penha-Goncalves, M. U. Ferreira, R. Aguas, Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. medRxiv 10.1101/2020.04.27.20081893 [Preprint]. 21 May 2020; https://doi.org/10.1101/2020.04.27.20081893.

14. F. Ball, T. Britton, P. Trapman, Code for: Population heterogeneity and consequences for herd immunity to SARS-CoV-2, Version 1, Zenodo (2020); https://doi.org/10.5281/zenodo.3899252.

15. H. Andersson, T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis* (Springer, 2000).

16. S. N. Ethier, T. G. Kurtz, *Markov Processes: Characterization and Convergence* (Wiley, 2009).