Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold

Authors: M. Gabriela M. Gomes[1,2*], Ricardo Aguas[3], Rodrigo M. Corder[4], Jessica G. King[5], Kate E. Langwig[6], Caetano Souto-Maior[7], Jorge Carneiro[8], Marcelo U. Ferreira[4], Carlos Penha-Gonçalves[8]

[1] Liverpool School of Tropical Medicine, Liverpool, United Kingdom.

[2] Centro de Matemática da Universidade do Porto, Porto, Portugal.

[3] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[4] Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil.

[5] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom.

[6] Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA.

[7] Laboratory of Systems Genetics, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA.

[8] Instituto Gulbenkian de Ciência, Oeiras, Portugal.

[*] Correspondence to: gabriela.gomes@strath.ac.uk.

Abstract

As severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spreads, the susceptible subpopulation is depleted causing the rate at which new cases occur to decline. Variation in individual susceptibility or exposure to infection exacerbates this effect. Individuals that are frailer, and therefore more susceptible or more exposed, have higher probabilities of being infected, depleting the susceptible subpopulation of those who are at higher risk of infection, and thus intensifying the deceleration in occurrence of new cases. Eventually, susceptible numbers become low enough to prevent epidemic growth or, in other words, herd immunity is attained. Although estimates vary, it is currently believed that herd immunity to SARS-CoV-2 requires 60-70% of the population to be immune. Here we show that variation in susceptibility or exposure to infection can reduce these estimates. Achieving accurate estimates of heterogeneity for SARS-CoV-2 is therefore of paramount importance in controlling the COVID-19 pandemic.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in China in late 2019 and spread worldwide causing the ongoing pandemic of coronavirus disease (COVID-19). As of 26 April 2020, more than 2.8 million cases have been confirmed and almost 200,000 died (*1*). Scientists throughout the world have engaged with governments, health agencies, and with each other, to address this emergency. Mathematical models, in particular, have been core to important decisions concerning contact tracing, quarantine, and social distancing, to mitigate or suppress the initial pandemic spread (*2*). Successful suppression, however, leaves populations at risk to resurgent waves due to insufficient acquisition of immunity. Models have thus also addressed longer term SARS-CoV-2 transmission scenarios and the requirements for continued adequate response (*3*). This is especially timely as countries begin to relax lockdown measures that have been in place over recent weeks with varying levels of success in tackling national outbreaks.

Here we demonstrate that individual variation in susceptibility or exposure (connectivity) accelerates the acquisition of immunity in populations. More susceptible and more connected individuals have a higher propensity to be infected and thus are likely to become immune earlier. Due to this *selective immunisation*, heterogeneous populations require less infections to cross their herd immunity thresholds than homogeneous (or not sufficiently heterogeneous) models would suggest. We integrate continuous distributions of susceptibility or connectivity in otherwise basic epidemic models for COVID-19 and show that as the coefficient of variation increases from 0 to 4, the herd immunity threshold declines from over 60% to less than 10%. Measures of individual variation are urgently needed to narrow the estimated ranges of herd immunity thresholds and plan accordingly.

SARS-CoV-2 transmission in heterogeneous populations

SARS-CoV-2 is transmitted primarily by respiratory droplets and modelled as a susceptible-exposed-infectious-recovered (SEIR) process.

Variation in susceptibility to infection

Individual variation in susceptibility is integrated as a continuously distributed factor that multiplies the force of infection upon individuals:

$$\dot{S}(x) = -\lambda x S(x), \quad \dot{E}(x) = \lambda x S(x) - \delta E(x), \quad \dot{I}(x) = \delta E(x) - \gamma I(x), \tag{1}$$

where $S(x)$ is the number of individuals with susceptibility x, $E(x)$ and $I(x)$ are the numbers of individuals who originally had susceptibility x and became exposed and infectious, δ is the rate of progression from exposed to infectious, γ is the rate of recovery or death, and $\lambda = (\beta/N)\int[\rho E(x) + I(x)]\,dx$ is the average force of infection upon susceptible individuals in a population of size N. The basic reproduction number for system (1) is:

$$R_0 = \langle x \rangle (\beta/N)(\rho/\delta + 1/\gamma), \tag{2}$$

where ρ is a factor representing the infectivity of individuals in compartment E in relation to those in I, and $\langle x \rangle$ is the mean susceptibility factor at epidemic onset. Prior to the epidemic, susceptibility is described by a probability density function $q(x)$ with mean 1 and coefficient of variation (CV) to be explored as a parameter.

Figure 1 depicts model (1) trajectories fitted to supressed epidemics in Italy and Austria (orange), assuming coefficients of variation 1 and 3. The corresponding uncontained scenarios are shown in black. The difference in epidemic sizes between the two levels of variation is

substantial. In the case of Italy, where suppression was less successful, the pandemic appears mostly resolved when $CV = 3$. However, a large second wave (or a series of smaller waves, depending on containment strategies) remains in the horizon when $CV = 1$. Countries where suppression of the initial outbreak was more successful, such as Austria, have acquired less immunity and therefore the potential for future transmission in the respective populations remains naturally larger. However, also in these situations, expectations for the potential of subsequent waves is much reduced by variation in susceptibility to infection.

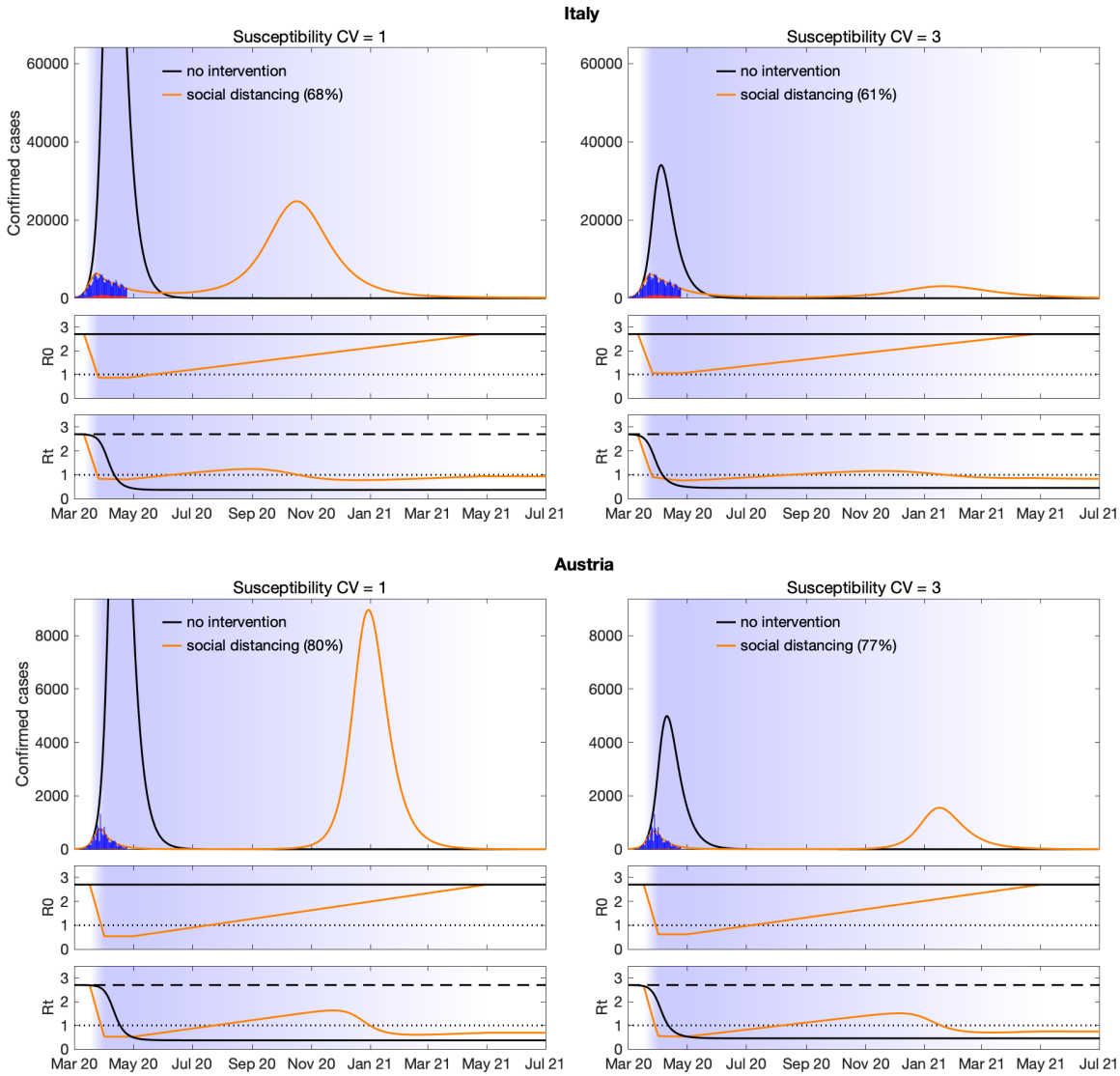

Figure 1: The effect of variation in susceptibility to infection on the size of epidemics. An uncontained epidemic (black) and an epidemic supressed by social distancing (orange). Blue bars are confirmed cases and overlaid red bars represent deaths. Basic (R_0) and effective (R_t) reproduction numbers are displayed on bottom panels. Shades represent social distancing periods: maximal over periods of one month between March and April 2020, with a prior 2-week ramp-up period and followed by a 1-year ramp-down. Susceptibility factors were implemented as gamma distributions. Parameter values: $\delta = 1/4$ per day; $\gamma = 1/4$ per day; $\rho = 0.5$; $R_0 = 2.7$. Fraction of infected individuals identified as positive (reporting fraction): $p = 0.1$.

Variation in exposure to infection

In a directly transmitted infectious disease, such as COVID-19, variation in exposure to infection is primarily governed by patterns of connectivity among individuals. We implement this in system (1) by adding variation in infectivity and assuming a positive correlation between susceptibility and infectivity. Formally this corresponds to modifying the force of infection as $\lambda = (\beta/N)(\int x[\rho E(x) + I(x)]\,dx / \int xq(x)\,dx)$ and the basic reproduction number as:

$$R_0 = (\langle x^2 \rangle / \langle x \rangle)(\beta/N)(\rho/\delta + 1/\gamma), \tag{3}$$

where $\langle x \rangle$ and $\langle x^2 \rangle$ are the first and second moments of the distribution $q(x)$ prior to the epidemic.

Applying this model to the epidemics in Italy and Austria (Figure 2) leads to similar results to those obtained when variation was in susceptibility to infection.

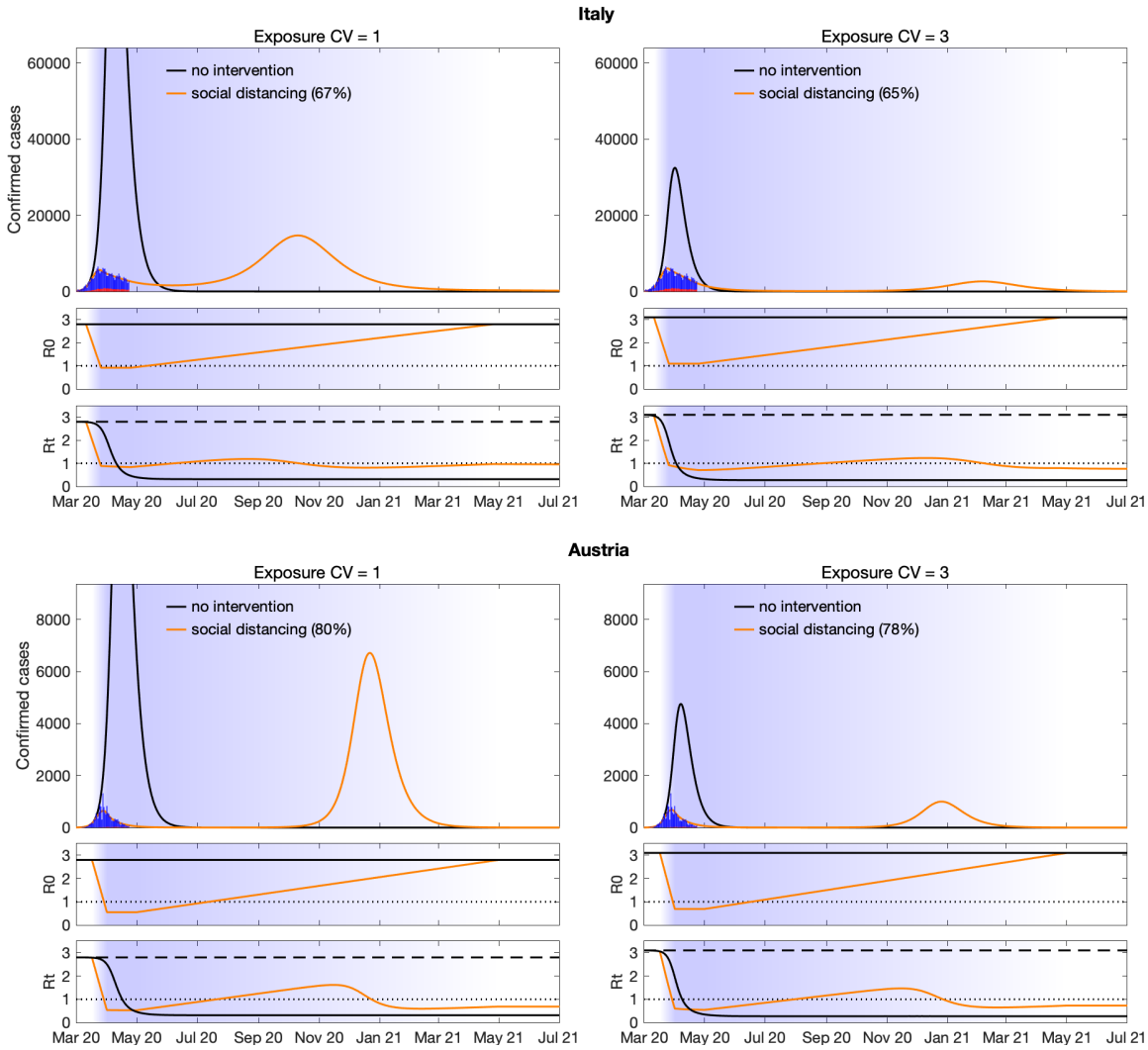

Figure 2: The effect of variation in connectivity to infection on the size of epidemics. An uncontained epidemic (black) and an epidemic supressed by social distancing (orange). Blue bars are confirmed cases and overlaid red bars represent deaths. Basic (R_0) and effective (R_t) reproduction numbers are displayed on bottom panels. Shades represent social distancing periods: maximal over periods of one month between March and April 2020, with a prior 2-week ramp-up period and followed by a 1-year ramp-down. Connectivity factors were implemented as gamma distributions. Parameter values: $\delta = 1/4$ per day; $\gamma = 1/4$ per day; $\rho = 0.5$; $R_0 = 2.8$ when $CV = 1$ and $R_0 = 3.1$ when $CV = 3$. Fraction of infected individuals identified as positive (reporting fraction): $p = 0.1$.

The herd immunity threshold

Individual variation in risk of acquiring infection is under selection by the force of infection, whether individual differences are due to biological susceptibility, physical exposure, or a combination of the two traits. Selection results in the removal of the most at-risk individuals from the susceptible pool as they become infected and eventually recover (some die). This selective acquisition of infection and immunity results simultaneously in decelerated epidemic growth and accelerated induction of immunity in the population. The herd immunity threshold defines the percentage of the population that needs to be immune to curve down epidemics and prevent future waves. Figure 3 shows the downward trends in the herd immunity threshold for SARS-CoV-2 as coefficients of variation for susceptibility and exposure are increased between 0 and 4. While herd immunity is expected to require 60-70% of a homogeneous population to be immune given an R_0 between 2.5 and 3 (*2*, *3*), these percentages drop to the range 10-20% for CVs between 2 and 4. Therefore, a critically important question is: how variable are humans in their susceptibility and exposure to SARS-CoV-2?

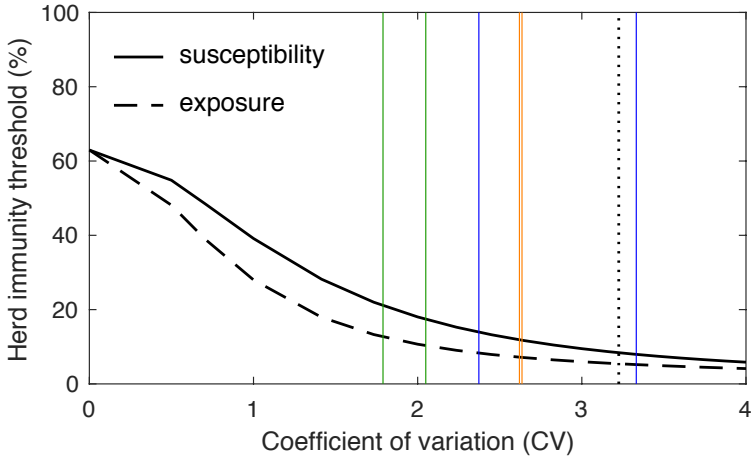

Figure 3: Herd immunity threshold with variation in susceptibility and exposure to infection. Vertical lines indicate coefficients of individual variation for several infectious diseases according to literature: (green) susceptibility or exposure to malaria [Amazon 1.8 (*4*), Africa 2.4 (*5*)]; (blue) susceptibility or exposure to tuberculosis [Portugal 2.4, Brazil 3.3 (*6*)]; (orange) infectiousness for SARS-CoV-1 [Singapore 2.62, Beijing 2.64 (*7*)]; (dotted black) infectiousness for SARS-CoV-2 [3.2 (*8*)].

As the pandemic unfolds evidence will accumulate in support of low or high coefficients of variation, but soon it will be too late to impact public health strategies. We searched the literature for estimates of individual variation in propensity to acquire or transmit COVID-19 or other infectious diseases and overlaid the findings as vertical lines in Figure 3. Most CV estimates are comprised between 2 and 4, a range where naturally acquired immunity to SARS-CoV-2 may place populations over the herd immunity threshold once as few as 10-20% of its individuals are immune. This depends, however, on which specific transmission traits are variable.

Variation in infectiousness was critical to attribute the scarce and explosive outbreaks to superspreaders when the SARS-CoV-1 emerged in 2002 (*7*), but infectiousness does not respond to selection as susceptibility or exposure do. Models with or without individual variation in infectiousness perform equivalently when implemented deterministically (Figure S1) and only differ due to stochasticity in an entirely different phenomenon to that presented in this paper (*8*, *9*). Among the estimates of individual variation plotted in Figure 3, those corresponding to

SARS-CoV viruses have been described as variation in individual infectiousness (*7*, *8*), but the way authors describe superspreaders is suggestive that higher infectiousness stems from higher connectivity with other individuals, who may be susceptible. This would support the scenarios displayed in Figure 2, with $CV = 3$ for exposure to infection.

Discussion

The concept of *herd immunity* is most commonly used in the design of vaccination programmes (*10*). Defining the percentage of the population that must be immune to cause infection incidences to decline, herd immunity thresholds constitute convenient targets for vaccination coverage. In idealised scenarios of vaccines delivered at random and individuals mixing at random, herd immunity thresholds are given by a simple formula $(1 - 1/R_0)$ which, in the case of SARS-CoV-2, suggests that 60-70% of the population should be immunised to halt spread considering estimates of R_0 between 2.5 and 3. A crucial caveat in exporting these calculations to immunization by natural infection, is that natural infection does not occur at random. Individuals who are more susceptible or more exposed are more prone to be infected and become immune earlier, which lowers the threshold. The herd immunity threshold declines sharply when coefficients of variation increase from 0 to 2 and remains below 20% for more variable populations.

Heterogeneity in the transmission of respiratory infections has traditionally focused on variation in exposure summarised into age-structured contact matrices (*11*, *12*). Besides overlooking differences in susceptibility given exposure, the aggregation of individuals into age groups curtails coefficients of variation with important downstream implications. Popular models based on contact matrices use a coefficient of variation around 0.9 (*13*) and perform similarly to our scenarios for $CV = 1$. Supported by existing estimates across infectious diseases, we argue that CV is generally higher and prognostics more optimistic than currently assumed. However plausible, this needs to be confirmed for the current COVID-19 pandemic and, given its relevance to policy decisions, it should be set as a priority.

Interventions themselves have potential to manipulate individual variation. Current social distancing measures may be argued to either increase or decrease variation in exposure, depending on the compliance of highly-susceptible or highly-connected individuals in relation to the average. A deep understanding of these patterns is crucial not only to develop more accurate predictive models, but also to refine control strategies and to interpret data resulting from ongoing serological surveys.

Based on the spectrum of current knowledge, a high level of pragmatism may be required in policy responses to serological surveys. On the one hand, if CV is very low, the most stringent control measures would need to be continued for suppression of the epidemic. The other side of that coin is a scenario where keeping only the mildest control measures (protecting the elderly to reduce mortality rates) is optimal. It would therefore be imperative to conduct longitudinal serological studies in representative samples of the population, as control measures are relaxed. Given a percent positivity in an initial survey, the speed at which that figure increases after control measures are eased would reveal what the most likely value of CV is, and simultaneously advise which control measures should be enforced.

References

1. World Health Organization, Coronavirus disease 2019 (COVID-19) Situation Report – 97 (WHO, 2020).

2. 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, Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand (Imperial College COVID-19 Response Team, 2020). 10.25561/77482.

3. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 10.1126/science.abb5793 (2020).

4. R. M. Corder, M. U. Ferreira, M. G. M. Gomes, Modelling the epidemiology of residual Plasmodium vivax malaria in a heterogeneous host population: a case study in the Amazon Basin. *PLOS Comput. Biol.* **16**, e1007377 (2020).

5. D. L. Smith, J. Dushoff, R. W. Snow, S. I. Hay, The entomological inoculation rate and its relation to the prevalence of Plasmodium falciparum infection in African children. *Nature* **438**, 492-495 (2005).

6. M. G. M. Gomes, J. F. Oliveira, A. Bertolde, D. Ayabina, T. A. Nguyen, E. L. Maciel, R. Duarte, B. H. Nguyen, P. B. Shete, C. Lienhardt C, Introducing risk inequality metrics in tuberculosis policy development. *Nat. Commun.* **10**, 2480 (2019).

7. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355-359 (2005).

8. A. Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, S. Abbott, A. J. Kucharski, S. Funk, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outsize China [version 1; peer review: awaiting peer review]. *Wellcome Open Res.* **5**, 76 (2020).

9. C. Donnat, S. Holmes, Modeling the heterogeneity in COVID-19's reproduction number and its impact on predictive scenarios. arXiv:2004.05272 (2020).

10. P. Fine, K. Eames, D. L. Heymann, "Herd immunity": a rough guide, *Clin. Infect. Dis.* **52**, 911-916 (2011).

11. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Med*. **5**, e74 (2008).

12. P. Klepac, A. J. Kucharski, A. J. K. Conlan, S. Kissler, M. Tang, H. Fry, J. R. Gog, Contacts in context: large-scale setting specific social mixing matrices from the BBC Pandemic project. medRxiv preprint: 10.1101/2020.02.16.20023754 (2020).

13. K. Prem. A. R. Cook, M. Jit, Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Comput. Biol.* **13**: e1005697 (2017).