Herd immunity in the epidemiology and control of COVID-19

This rapid review of the science of herd immunity and COVID-19 from the Royal Society is provided to assist SAGE in relation to COVID-19.

This paper is a pre-print and has been subject to formal peer-review.

Summary
When a sufficiently large fraction of people are immune to an infectious disease, the entire population can be protected by “herd immunity”. But herd immunity acquired by natural infection has several disadvantages as a means of COVID-19 control: (1) the build-up of herd immunity would be associated with a high burden of illness and death; (2) while infection is spreading through a population, it is not yet clear that the most vulnerable can be protected from severe and fatal COVID-19; and (3) it may not be possible to achieve herd immunity by natural infection if protection against reinfection is partial and transient. The preferred route to herd immunity is not through natural infection but by vaccination. While awaiting population-wide coverage of COVID-19 vaccines, the principal method of control is to prevent transmission between infectious and uninfected people using non-pharmaceutical interventions, i.e. antiviral hygiene, physical barriers and personal distancing. A large number of people are still at risk of serious illness and death from COVID-19 in most countries around the world; consequently, it remains critical to reduce infectious contacts below the level at which transmission can be sustained, i.e. by keeping the case reproduction number below one.

Introduction
When a sufficiently large fraction of people are immune to an infectious disease – a fraction above the herd immunity threshold – the entire population can be protected from an epidemic. In principle, herd immunity could be established naturally by infection or artificially by vaccination. The purpose of this paper is to explain the circumstances under which herd immunity can and cannot protect individuals and populations from SARS CoV-2 infection and from COVID-19 disease.

Herd immunity in principle
When a transmissible pathogen such as SARS CoV-2 invades a previously unexposed population, the primary (index) case generates an average of $R_0$ infectious secondary cases. $R_0$ is the basic case reproduction number and the secondary cases arise over the generation time of the pathogen ($\approx 5$ days on average for SARS CoV-2). If $R_0 > 1$ infection spreads; if $R_0 \leq 1$, infection dies out.

As an epidemic grows over time, $t$, so does the frequency of encounters between those people who are infectious and those who have previously been exposed. If exposure to infection stimulates protective immunity, then the generation case reproduction number falls to $R_t = (1 - p_t)R_0$, where $p_t$ is the proportion of people immune to infection at time $t$. Immunity here refers to protection from a further episode of infectiousness, with the potential for transmission, and not necessarily protection from disease.
Any level of immunity in the population reduces the case reproduction number below its (average) maximum value of $R_0$ (Figure 1a, black line), so the growth of the epidemic slows as infection spreads (Figure 1a, grey line). When $p_t$ reaches a value of $1 - 1/R_0$ then $R_t = 1$. This is the herd immunity threshold. As soon as $p_t$ exceeds that threshold the incidence of disease begins to fall (Figure 1a, grey line) but transmission continues until the proportion of people infected reaches a maximum of $p_f = 1 - e^{-p_t R_0}$, which is the “final size” of an uncontrolled epidemic (Figure 1a, dashed line).

Because the final size of an epidemic is necessarily greater than the herd immunity threshold, the whole population is protected from a subsequent epidemic until the proportion of people immune subsides to the threshold value of $1 - 1/R_0$, either by the loss of individual immunity or by the introduction of new susceptibles through birth or immigration.

For SARS CoV-2, estimates of $R_0$ are typically in the range 2 to 4, which is larger, for example, than the values for influenza A and Ebola but smaller than those for polio and rubella (Figure 1b). If $R_0 = 3$, the herd immunity threshold is 0.66 and the final size of the epidemic is 0.94. Thus, in an uncontrolled SARS CoV-2 epidemic, almost everyone would be infected. Naturally acquired herd immunity protects very few people during the first epidemic wave – only 6% of the population remains uninfected before infection dies out.

The value of herd immunity acquired during the first epidemic is in preventing a second epidemic. If the 94% of people exposed in the first epidemic have complete and lasting immunity to reinfection, then the whole population would be protected. The percentage immune is far above the herd immunity threshold (94% > 66%) and, in this deterministic formulation of the theory, a subsequent outbreak would be impossible.

This simple theory captures the essence of herd immunity, but reality is more complex. Three elaborations are worth mentioning.

First, herd immunity does not make epidemics impossible, only improbable. The basic case reproduction number, $R_0$, summarizes a series of events that are probabilistic (stochastic) rather than deterministic, such as the probability of transmission when two people meet, or the likelihood that a COVID-19 patient remains infectious from one day to the next. By chance, any of these events might happen more or less often than expected. Consequently, the threshold at $R_0 \approx 1$ (actually a little bigger than 1 in a world of chance) is not a boundary between the absence or presence of infection in a population; rather, it separates the risk of smaller local outbreaks from larger population-wide epidemics.

**FIGURE 1**

a) Build-up of herd immunity during a model epidemic of SARS CoV-2.

b) Herd immunity threshold in relation to $R_0$, comparing SARS CoV-2 with influenza A, Ebola virus disease, poliomyelitis and rubella

Sources: epidemic model and data.134.
Second, simple theory presents $R_0$ as an average for a homogeneously mixing population. When mixing between infectious and susceptible people is not homogeneous - for example, when there are different rates of contact within population sub-groups - the herd immunity threshold for the whole population is lower than under homogeneous mixing, i.e. less than $(1 - 1/R_0)$. One illustrative study of variable contact rates within and between people in different age groups showed how a herd immunity threshold of 60% (for average $R_0 = 2.5$) could be reduced to 44%\(^6\). This example reveals the general effect of non-homogeneous mixing in structured populations, but the herd immunity threshold would vary from one setting to another.

There is one further consequence of above-average $R_0$: if an outbreak is started by an infectious super-spreader, or by an infectious person at a super-spreading event (e.g. intense social mixing), a population at the herd immunity threshold (based on average $R_0$) will not be protected from (at least) a small outbreak of disease. Super-spreading has frequently been reported for SARS CoV-2, as for example in China and Hong Kong\(^7\). An uneven or overdispersed distribution of secondary infections from primary transmitters is expected to be the norm for SARS CoV-2.

Third, if infection does not lead to complete and lasting immunity, then $(1 - 1/R_0)$ is an underestimate of the herd immunity threshold. If, for example, the proportion of people who acquire immunity from a SARS CoV-2 infection is a fraction $\varepsilon$ (so that a proportion of people 1- $\varepsilon$ remain completely susceptible to reinfection), then the herd immunity threshold would have the larger value of $(1 - 1/R_0)/\varepsilon$. This criterion applies whether immunity is acquired by natural infection or by vaccination\(^8\).

These effects of probability, heterogeneity and partial immunity indicate that the herd immunity threshold $(1 - 1/R_0)$ should be considered a minimum requirement for population protection. From the perspective of epidemiology alone, putting aside other considerations such as economics, the higher the fraction of people immune, the lower the chance of an outbreak of infection. It is worth stressing, though, that anyone who acquires immunity to infection is not only protected personally but is also prevented from transmitting infection to others.

The main alternative to protecting a population by herd immunity is to prevent transmission between infectious and susceptible individuals. For any infectious disease, the value of $R_0$ is determined by the average duration of infectiousness (measured e.g. in days) multiplied by the number of contacts made by an infectious person each day. For SARS CoV-2, there is presently no way of reducing the duration of infectiousness (e.g. antiviral drugs are used only to support clinical care), but it is possible to reduce transmission with a range of “non-pharmaceutical interventions” (NPIs), including antiviral hygiene, physical barriers and personal distancing. The criterion for population protection by NPIs is similar to that for herd immunity: if the contact rate is reduced by a factor of $(1 - 1/R_0)$ or more, then $R_0 \leq 1$ and infection cannot be established or maintained in a population. The drawback is that, with little or no build-up of herd immunity, the population remains vulnerable to outbreaks if and when NPIs are lifted.

The strategy of reducing person-to-person transmission by NPIs has been pursued relatively successfully in China, Germany, Hong Kong, New Zealand, Republic of Korea, Singapore and Taiwan, among other countries. Few countries have considered protecting their populations by allowing the spread of infection to create herd immunity\(^9\). However, in Manaus, Brazil, an estimated three quarters of the population were infected in the first, largely unmitigated epidemic wave. Manaus is perhaps the only large city in the world to have exceeded the putative herd immunity threshold for SARS CoV-2, but it remains to be seen whether that population is protected from further outbreaks of COVID-19\(^9\).

**Herd immunity in practice**

Given the principles outlined above, is it possible to protect populations from SARS CoV-2 and COVID-19 by herd immunity? There are three inter-related problems in achieving herd immunity by means of natural infection, and in protecting populations from second and subsequent waves of infection and disease.

1. **The route to herd immunity through natural infection carries a large penalty in illness and death**

As explained above, naturally acquired herd immunity would protect almost no-one during a first epidemic wave, even if infection is allowed to spread without control. In reality, no epidemic of SARS CoV-2 has been uncontrolled. Lockdown in the UK from 23 March onwards, employing an array of NPIs, reduced the basic case reproduction number, $R_0$, from an estimated 3.0 - 4.0 down to a value in the range 0.44 - 0.82 through April 2020\(^10\). Consequently, the proportion of people infected in the first epidemic wave was lower than the herd immunity threshold, and far lower than the expected maximum size of the epidemic.

Approximately 40,000 people died with COVID-19 during the first epidemic wave in the UK (1 in 1666 people, a conservative estimate). With an infection fatality ratio of the order of 1% (1.04%, range 0.84 - 1.19% by one estimate for England)\(^12\), these deaths would have occurred among roughly 4 million infected people, or 6% of the whole UK population.
Lockdown evidently saved many lives during the first epidemic wave in the UK (and earlier lockdown would have saved more): given the potential for epidemic spread ($R_0 = 3 - 4$), the price of attempting to achieve herd immunity in an unmitigated first wave of COVID-19 would almost certainly have been hundreds of thousands of deaths, and many more episodes of serious illness, overwhelming health and social services, and damaging lives and livelihoods\(^1\). Given that most people have still not been infected in the UK, the further build-up of immunity during a second epidemic wave would be associated with many more episodes of illness and many more deaths. The current (October - November 2020) resurgence of COVID-19 supports this view.

2. It is not yet possible to identify and protect all vulnerable people from severe forms of COVID-19

It has been proposed that infection and herd immunity in less vulnerable members of a population (e.g. children and young adults) could protect those who are more vulnerable\(^4\). The first challenge in shielding the vulnerable is that there are a great number and diversity of people at relatively high risk of severe COVID-19, through a combination of exposure to infection and susceptibility to disease. People at higher risk include those in older age groups, those vulnerable by ethnic and social group (e.g. black and minority ethnic people, people in low-income groups), or those with underlying health conditions (e.g. diabetes, obesity or immunosuppressive diseases or medical treatments)\(^1\). Older people illustrate the practical difficulties of shielding. The mortality rate among people with COVID-19 increases in parallel with the death rate from all causes, so the death rate among COVID-19 patients over 90 years old is thousands of times higher than for children under 15. The task then is to choose a threshold age for shielding – protecting more younger people at lower risk, or fewer older people at higher risk\(^4\). Despite the logic and variety of proposals for shielding older people from infection, none has yet been unambiguously effective in practice\(^4\). Routinely collected data on COVID-19 deaths by age underline the point. In the United States and in European countries, the spread of coronavirus infection among younger people has leaked into older age groups\(^4\). During the first epidemic wave in the UK (March - April 2020), the number of deaths in people aged 60 years and over was far greater than in younger age groups (Figure 2a). The second wave of infections in the UK (October 2020) grew more slowly, but the death rate of older people was still comparatively high (Figure 2b). Besides being ineffective, the separation of people who are both vulnerable and dependent from the rest of society is also unpopular.

Although the characteristics of people at high risk of COVID-19 are becoming clearer, there is no group of people for whom severe COVID-19 is known to be risk-free. The emerging phenomenon of “Long Covid” makes the point that targeted shielding could not...
protect everyone at risk of serious illness\textsuperscript{17,18}. Long Covid is multifarious: patients with “ongoing symptomatic COVID-19” (4 - 12 weeks after the initial onset of illness) and “post-COVID-19 syndrome” (>12 weeks) show manifold signs and symptoms, which can arise from any system in the body, which often overlap, and which change over time. The symptoms include severe fatigue, reduced exercise capacity, breathlessness, chest pain, fever, palpitations, anxiety and cognitive impairment, and loss of smell and taste. They reflect the involvement of heart, lung, gut, brain, blood, skin, sensory and nervous systems, among others\textsuperscript{18-21}. The aetiology and epidemiology of Long Covid are still under investigation, but around one quarter to one third of COVID-19 patients under the age of 65 years suffer prolonged illness\textsuperscript{22-26}. There are no specific treatments for Long Covid, apart from the management of symptoms. As details continue to emerge about this syndrome, it is already clear that Long Covid would be an additional, substantial and unpreventable cost of permitting widespread infection in order to achieve herd immunity.

3. Acquired immunity to SARS CoV-2 infection may be partial and transient
An uncontrolled SARS CoV-2 epidemic would infect the majority of people in any community (Figure 1a). But population immunity is not guaranteed if individual immunity is partial or transient. Protective immunity to the other four endemic human coronaviruses (besides SARS CoV-2) wanes after 1 to 3 years, allowing frequent re-infection. Humoral (antibodies produced by B cells) and cellular (T cells) immune responses to SARS CoV-2 infection also decline on a timescale of weeks or months\textsuperscript{27-32}. These observations suggest that protection against re-infection might also wane on this timescale; on the other hand, several preliminary reports indicate that infection stimulates a protective immune response for longer than few weeks, perhaps many months or years\textsuperscript{39,41}. Besides the waning of immunity in human hosts, viral mutations might also permit SARS CoV-2 to evade immunity while continuing to be infective and transmissible\textsuperscript{33}. The emergence of such SARS CoV-2 genetic variants would also limit the efficacy and duration of immunity conferred by a primary SARS CoV-2 infection, reducing the level of protection in individuals and populations.

In addition, a growing number of studies are reporting re-infection with SARS CoV-2, confirmed by the isolation of two genetically distinctly viruses from the same patient\textsuperscript{34}. The frequency of re-infection and its implications for individual and population immunity are not yet known, but these findings are consistent with the broad picture of incomplete and/or transient immunity to SARS CoV-2. In the context of this paper, protective immunity in individuals is important because it is the basis of herd immunity, which influences the dynamics of infection in populations. If there is widespread exposure to infection, and individual immunity is complete and lifelong, we expect to see periodic outbreaks of infection as the population of susceptibles is renewed by births, with inter-epidemic periods of several years. Epidemiological patterns of this kind were familiar for other immunizing viral diseases such as measles, mumps and rubella, prior to widespread vaccination\textsuperscript{35}. On the other hand, if immunity is partial and transient, we expect to see persistent endemic infection with damped or low amplitude oscillations and shorter inter-epidemic periods. The dynamics of SARS CoV-2 would also be modulated by seasonal variation in transmission, by cross-immunity to other coronaviruses, by competition with other co-circulating respiratory viruses, and by vaccination\textsuperscript{36-38}. While the various epidemiological possibilities have been mapped out in theory, it is not yet clear what will happen in practice. What is certain is that any form of protective immunity in individuals and populations will influence the dynamics of SARS CoV-2 for years to come. All of the scenarios imagined by modelling suppose that COVID-19 will persist as an endemic disease and take its place alongside the other four seasonal coronaviruses. The elimination of SARS CoV-2 will only be possible if vaccines, with high efficacy and safety and with wide coverage, can keep populations above the herd immunity threshold\textsuperscript{8,37}. It is unlikely that this coronavirus can be eliminated by non-pharmaceutical interventions alone.
Conclusion

Naturally acquired herd immunity might protect populations from SARS CoV-2 infection, at least partially and temporarily, but has several disadvantages as a means of COVID-19 control:

• Herd immunity to SARS CoV-2 protects no more than a small fraction of a population from infection during a first epidemic wave, whatever proportion of people become immune following infection.

• The development of herd immunity by allowing widespread infection would be associated with a high burden of illness and death, far higher than already reported.

• It is not yet possible to protect all those who are vulnerable to COVID-19, either from infection or disease; nor is it feasible to achieve herd immunity in people at low risk of severe COVID-19 in order to protect those at high risk.

• It may not be possible to sustain the herd immunity threshold by natural infection if immunity to reinfection is partial and/or transient.

Herd immunity could contribute to COVID-19 control if these inferences about infection, immunity and illness are wrong or change, i.e. if immunity proves to be highly protective and lasts for years; if vulnerable people can be shielded from serious illness and death; or if cheap and effective COVID-19 treatments become available. These factors therefore need to be kept under review, and expectations about the role of population immunity in COVID-19 epidemiology and control adjusted accordingly.

In any event, the preferred route to herd immunity is through vaccination. Vaccination would protect each person immunized, and also prevent the transmission of infection to others. The first priority for a new SARS CoV-2 vaccine is to protect people who are at greatest risk of severe illness. However, if and when enough people can be immunized so as to exceed the herd immunity threshold, then the whole population would be protected too. Preliminary reports of high efficacy among candidate vaccines suggest that the most vulnerable can be protected and that herd immunity might be achievable.

At present, the principal method of controlling SARS CoV-2 infection and COVID-19 disease is by preventing transmission between infectious and uninfected people through non-pharmaceutical interventions (NPIs). Without any guarantee of long-lasting protection from individual or population immunity, and with the continuing risk of serious illness and death, there is still a high premium on reducing infectious contacts below the level at which transmission can be sustained, i.e. keeping the case reproduction number below one (R_t ≤ 1). But that premium has an upper limit: the economic, social and other health costs of curtailing coronavirus transmission will be set against the benefits of controlling COVID-19. The goal of COVID-19 control is not merely to be effective, but to be cost-effective too, protecting both lives and livelihoods.
References


40. Dan JMM et al. 2020 Immunological memory to SARS-CoV-2 assessed for greater than six months after infection. bioRxiv. (doi:2020.11.15.383323)


43. Voysey M et al. 2020 Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. The Lancet. (doi:10.1016/S0140-6736(20)32661-1)

44. Smith GD, Spiegelhalter D. 2020 Shielding from covid-19 should be stratified by risk. BMJ, m2063. (doi:10.1136/bmj.m2063)


**DISCLAIMER**

This paper has drawn on the most recent evidence up to 26 November 2020 and has been subject to formal peer-review. Further evidence on this topic is constantly published and the Royal Society may return to this topic in the future. This independent overview of the science has been provided in good faith by subject experts and the Royal Society and paper authors accept no legal liability for decisions made based on this evidence.

**THANKS**

The Royal Society is grateful to the Leverhulme Trust for its support for the Society’s pandemic response work.

The text of this work is licensed under the terms of the Creative Commons Attribution License which permits unrestricted use, provided the original author and source are credited. The license is available at: creativeworks.org/licenses/by/4.0

Issued: November 2020 DES7286 © The Royal Society