Article:
Ashby, B. and Best, A. (2021) Herd immunity. Current Biology, 31 (4). R174-R177. ISSN 0960-9822

https://doi.org/10.1016/j.cub.2021.01.006

Herd immunity

Ben Ashby[1,*] and Alex Best[2]

Herd immunity is an important yet often misunderstood concept in epidemiology. As immunity accumulates in a population — naturally during the course of an epidemic or through vaccination — the spread of infectious disease is limited by the depletion of susceptible hosts. If a sufficient proportion of the population is immune — above the 'herd immunity threshold' — then transmission generally cannot be sustained. Maintaining herd immunity is therefore critical to long-term disease control. In this primer, we discuss the concept of herd immunity from first principles, clarify common misconceptions, and consider the implications for disease control.

What is herd immunity?

The notion of herd immunity is simple, yet profound: not every member of a population must be immune to prevent large-scale outbreaks, nor will everyone be infected during the course of an epidemic (Figure 1). It is both a fundamental epidemiological concept describing a natural phenomenon, and a practical goal for long-term disease control, most commonly associated with vaccination programs. Recently, during the COVID-19 pandemic, the subject of herd immunity — specifically, how and when it might be achieved — has received considerable attention from scientists, policymakers and the general public. Yet despite its apparent simplicity, misconceptions about herd immunity and its implications for disease control are surprisingly common. Part of the confusion is due to the different ways in which immunity may be acquired (naturally through infection or by vaccination), but there is also significant variation in use of the term 'herd immunity', with some referring to whether a population has achieved a threshold level of immunity, and others to the extent of immunity

in the population (we will use ‘herd immunity’ in the threshold sense). For clarity, we begin by discussing the origins of the herd immunity concept.

Ideas about herd immunity first gained traction during the early 20th century, following experiments in mice by the bacteriologist William Topley and observations of diphtheria epidemics at the Royal Hospital School in Greenwich by Sheldon Dudley. These early insights were critical, as prior ideas about immunity focused almost entirely on the individual, neglecting population or ‘herd’ level effects. In 1927, these ideas were crystalised mathematically by two epidemiologists, William Kermack and Anderson McKendrick, who proposed the first compartmental models of infectious disease dynamics, which still form the cornerstone of epidemiological models today. By splitting a population into individuals who are susceptible, infected, or recovered, and considering the rates of movement between these classes, Kermack and McKendrick showed that epidemics typically produce a hump-shaped curve, encapsulating an exponential growth phase in infections, followed by a peak and subsequent decline in cases (Figure 2).

One can understand the principle of herd immunity by considering the different phases of an epidemic. When everyone in the population is susceptible to infection — for example, when a new pathogen enters a population that has no pre-existing immunity — a single infection produces R_0 new infections, on average. The quantity R_0 is known as the ‘basic reproduction number’ and is very important as it tells us whether the number of infections may initially grow ($R_0 > 1$) or will decline ($R_0 < 1$). Conceptually, R_0 can be understood as the product of several average quantities: the number of susceptible contacts for an infectious individual per unit time (or the contact rate), the transmission probability per contact, and the duration of infectiousness. Crucially, this means R_0 is not a fixed quantity and may vary between populations or over time.

Assuming the population is homogeneous (i.e. no variation in contact rates) and mixing is random, during the course of an epidemic each infection produces an average of $R = R_0 S(1 - i)$ new cases (R is a more general 'reproduction number'), where S is the proportion of the population that is susceptible and i is the relative strength of non-pharmaceutical interventions (NPIs) such as social distancing. It follows that cases will grow when $R > 1$ and decline when $R < 1$, which may occur due to interventions ($0 < i \leq 1$) or as the pool of susceptible individuals is depleted. Assuming immunity accumulates in the population, R naturally decreases even in the absence of NPIs ($i = 0$), with the epidemic peaking when $R = 1$. We can therefore deduce the threshold for herd immunity, h, by setting $R = 1$, $i = 0$, and $S = 1 - h$ in the equation above, and rearranging to give $h = 1 - \frac{1}{R_0}$. We therefore only need to know R_0 to determine the threshold for herd immunity, which can be readily calculated from epidemiological data. For example, the R_0 for COVID-19 is estimated to be around 2–4 (depending on the population and the variant) and for measles is in the range 12–18. These give approximate herd immunity thresholds of 50–75% and 92–94%, respectively (Figure 3). Intuitively, higher values of R_0 — due to greater transmissibility, higher contact rates, or longer infectious periods — correspond to higher thresholds for herd immunity.

This result has a number of profound implications. First, it tells us that in the absence of interventions (including behaviour change) a population naturally reaches herd immunity when the epidemic peaks ($R = 1$). Second, the threshold is equivalent to the level of vaccination that must be maintained in the population to prevent an epidemic: if a proportion, $p > h$, of the population is successfully vaccinated, then $R < 1$. Finally, the threshold is independent of the number of infections in the population. The model is agnostic to whether immunity is acquired naturally or through vaccination, hence the threshold for herd immunity is identical. However, the prevalence of infection when herd immunity is achieved may differ

greatly. In the case of rolling vaccination programs for diseases such as polio and measles, prevalence is typically very low. But if herd immunity is achieved naturally during the course of an epidemic, then this must occur when disease prevalence peaks. Reaching herd immunity therefore does not mean the end of an epidemic — a common misconception — as many individuals will continue to be infected while cases decline (Figure 2). For example, when $R_0 = 3$ the herd immunity threshold is 67%; vaccinating this proportion will prevent an epidemic. But if immunity is acquired naturally, then the epidemic will only peak when 67% have been infected and by the end of the epidemic 90% will have been infected. Achieving herd immunity simply means that $R < 1$ and so cases will decline from their current level, whether prevalence is high (naturally acquired) or low (vaccination).

Common misconceptions

The belief that herd immunity implies low or even zero disease prevalence is one of many misconceptions. Another relates to the indirect nature of protection conferred to the individual. Herd immunity reduces cases and therefore the likelihood of coming into contact with a pathogen, but susceptible individuals remain at risk of infection. Similarly, herd immunity prevents large-scale outbreaks from occurring because epidemic growth is unsustainable ($R < 1$), but infections may rise in the short-term. An important, but often overlooked principle of herd immunity is that it operates at a local level, and so the distribution of immunity in the population is crucial. The threshold is based on a well-mixed population with immunity randomly distributed, but if these assumptions do not hold then localised outbreaks may still occur even if the population as a whole is above the threshold.

The relationship between the herd immunity threshold and the peak of an epidemic can also lead to misinterpretations, because both occur when $R = 1$. However, the herd immunity threshold is calculated in the absence of interventions ($i = 0$). Interventions during an

epidemic ($i > 0$) will cause infections to peak before herd immunity has been reached. Therefore, one should not assume that herd immunity has been reached simply because an epidemic has peaked. Indeed, following the first waves of COVID-19, there were suggestions that many countries had achieved herd immunity since cases were in decline. However, serological testing revealed that relatively few people had been infected during the first wave (~5–10% in most cases). A resurgence of cases in late 2020 confirmed that herd immunity had not been reached. With no evidence of widespread reinfections or pre-existing immunity, the most parsimonious explanation is that public health interventions (NPIs) rather than herd immunity caused the epidemic peaks in early 2020.

Other prominent misconceptions relate to the dynamic, rather than fixed nature of herd immunity status. It is sometimes mistakenly claimed that we have never naturally achieved herd immunity to any pathogens, which appears to be due to the incorrect assumption that herd immunity implies elimination. Herd immunity is not a permanent state, and it may be temporarily achieved only to be lost through various processes, allowing pathogens to persist.

How is herd immunity lost?

A population that has achieved herd immunity may gradually or suddenly lose this status, for example, due to changes in population contact patterns. Alternatively, if the host immune response wanes over time, as is the case with pertussis (whooping cough), then the level of immunity in the population will steadily fall unless maintained through vaccination. When it falls below the herd immunity threshold, another epidemic may occur since $R > 1$, although the size of the epidemic will be much lower than in a completely susceptible population (Figure 4). Cases will decline once a sufficient number of individuals have been reinfected for the herd immunity threshold to be reached again. In principle, this cycle may repeat

indefinitely or with diminishing epidemic sizes until a stable endemic equilibrium is reached with $S = 1/R_0$ and $R = 1$.

Even if protection is lifelong, the level of immunity is gradually but inevitably eroded through population turnover. Immune individuals may die from other causes, while births lead to a steady influx of newly susceptible hosts (a net immigration of susceptible individuals has a similar effect). As with waning immunity, the population will likely experience repeated epidemic cycles in the absence of interventions (Figure 4). Hence, rolling vaccination programs exist for diseases such as measles, rubella, and polio, to maintain herd immunity.

Whereas waning immune responses and population turnover may lead to a gradual loss of immunity in the population, a sudden loss may occur due to pathogen evolution. If a new variant emerges with different antigens, then previously immune hosts may become susceptible. The recently discovered variant of COVID-19 in the UK (known as VUI 202012/01) is not thought to be sufficiently different to render previous immunity or vaccines ineffective, although it does appear to be significantly more transmissible. Antigenic evolution is especially common in RNA viruses due to their rapid mutation rates. For example, ‘antigenic drift’ in influenza viruses means that a new vaccination is required each year. This is distinct from booster vaccines which top-up waning immunity to the same antigens. Some influenza viruses may also undergo ‘antigenic shift’, where reassortment of different strains results in a novel phenotype to which the population may have little or no immunity (for example, the 2009 H1N1 ‘swine flu’ pandemic).

Implications for disease control

Herd immunity is critical for the long-term control of many infectious diseases. Since vaccines are never 100% effective and uptake is imperfect, achieving herd immunity offers a means of indirectly protecting those who remain at risk of infection, including those who are unable to be vaccinated due to their age or health. Although the herd immunity threshold is based on a simple model, it is remarkably accurate for informing vaccination programs. Herd immunity has been achieved by vaccination for a number of infectious diseases, leading to the global eradication of smallpox and rinderpest, with polio and several other diseases near eradication and many others heavily supressed or locally eliminated. However, until global eradication is achieved, countries that do not maintain vaccination above the herd immunity threshold may experience a resurgence.

A classic example is measles in the UK. Prior to a measles vaccine, there were between 200,000 and 800,000 reported cases of measles in the UK annually. With an estimated R_0 of 15 in the UK, the herd immunity threshold suggests 93% of the population should be vaccinated to prevent its spread. After the vaccine was introduced in 1971, uptake gradually rose and the number of cases quickly fell below 100,000 per year. In 1992, by which time the MMR (measles, mumps and rubella) combination vaccine had been introduced, the 93% threshold was reached. For the next 10 years measles cases never rose above 500 annually, a precipitous decline from the pre-vaccine era, putting the UK on the brink of eliminating measles. However, the false claims of a link between the MMR vaccine and autism in 1996 caused vaccine uptake to decline, reaching a nadir of 80% in 2004. There have since been multiple measles outbreaks, and although vaccine uptake has improved, the UK lost its elimination status from the World Health Organization in 2018.

Although the herd immunity threshold is identical whether it is achieved through vaccination or is naturally acquired through infection, the implications for disease control differ greatly. If a population were to pursue herd immunity naturally, then the overall disease burden must be based on the final epidemic size, which may be much higher than the herd immunity threshold. Furthermore, although a pathogen may be naturally eliminated at a local level due to a build-up of immunity in the population, the inevitable loss of herd immunity due to population turnover followed by reintroductions from other populations would likely lead to a resurgence in cases. Thus, naturally acquired herd immunity is not a viable long-term disease-control strategy.

During the COVID-19 pandemic, it was suggested that pursuing naturally acquired herd immunity would be preferable to socially and economically costly NPIs such as national lockdowns while vaccines were in development. Since the risk profile for mortality is heavily skewed towards the elderly and those with certain pre-existing conditions, in principle one could achieve herd immunity by shielding higher-risk individuals while allowing disease to spread among those at lower risk. Although theoretically possible, practically such an approach is not advisable for many reasons, including: the inability to effectively shield higher-risk individuals, especially people living in households with those at lower-risk; people may be poor judges of their own risk or may have undiagnosed co-morbidities; uneven distribution of immunity would likely lead to subsequent local outbreaks; if shielding fails then a long, strict lockdown will be required to bring cases under control; unnecessary mortality and morbidity (such as the so-called long-term COVID sequelae) among lower-risk individuals; potential for overwhelming healthcare capacity, leading to an increase in mortality from all causes; unknown duration or efficacy of naturally acquired immunity; increased mutation supply, leading to the emergence of new variants; and ethical implications for the prolonged isolation of higher-risk individuals with reduced access to health and social

care. One must compare these factors to the costs of NPIs and the pace of development of vaccines, along with their likely availability and efficacy. At the time of writing three vaccines have already shown efficacies of up to 90 or 95% and are expected to be widely distributed in 2021. Vaccination programs already underway in several countries and it is possible that herd immunity will be achieved later this year.

Conclusion

Usage of the term 'herd immunity' varies, but it is best reserved to describe the threshold phenomenon where a sufficient level of immunity in the population prevents epidemic growth ($R < 1$), and therefore populations either do, or do not, have herd immunity status. This status is not permanent, however, with population turnover among the factors that will lead to its loss. Herd immunity can be achieved naturally or by vaccination, yet there are important differences between the two mechanisms. The COVID-19 pandemic has elevated herd immunity from an uncontroversial concept to the focus of intense public debates, and although the principle is straightforward to describe, it is easy to misunderstand.

Further reading

Anderson, R.M., and May, R.M. (1985). Vaccination and herd immunity to infectious diseases. Nature *318*, 323–329.

Anderson, R.M., and May, R.M. (1991). Infectious diseases of humans: dynamics and control (Oxford, UK: Oxford University Press).

Brett, T.S., and Rohani, P. (2020). Transmission dynamics reveal the impracticality of COVID-19 herd immunity strategies. Proc. Natl. Acad. Sci. USA *117*, 25897–25903.

Britton, T., Ball, F., and Trapman, P. (2020). A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science. *369*, 846–849.

Jansen, V.A.A., Stollenwerk, N., Jensen, H.J., Ramsay, M.E., Edmunds, W.J., and Rhodes, C.J. (2003). Measles outbreaks in a population with declining vaccine uptake. Science. *301*, 804.

Kermack, W.O., and McKendrick, A.G. (1927) A Contribution to the Mathematical Theory of Epidemics. Proc. R. Soc. A *115*:700–721.

Thompson, R.N., Hollingsworth, T.D., Isham, V., Arribas-Bel, D., Ashby, B., Britton, T., Challenor, P., Chappell, L.H.K., Clapham, H., Cunniffe, N.J., *et al.* (2020). Key questions for modelling COVID-19 exit strategies. Proc. R. Soc. B *287*, 20201405.

[1]Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK. [2] School of Mathematics & Statistics, University of Sheffield, Sheffield, S3 7RH, UK.

*Email: benashbyevo@gmail.com

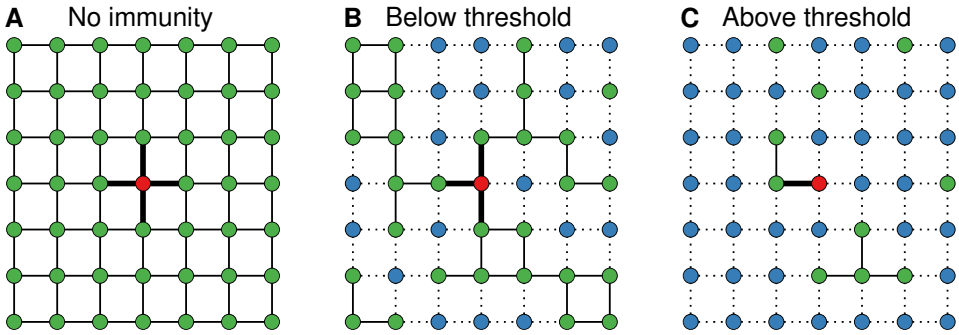

Figure 1. Herd immunity is achieved when a sufficient proportion of the population is immune to infection.

When there is no immunity in the population (A), an infectious individual (red) can readily spread disease to its contacts (bold lines), who are susceptible (green) and can transmit to their susceptible contacts (thin lines). When some individuals are immune (blue) but the population is below the herd immunity threshold (B), a large outbreak may still occur. When the population is above the threshold (C), large epidemics are prevented but small outbreaks may still occur among clusters of susceptible individuals.

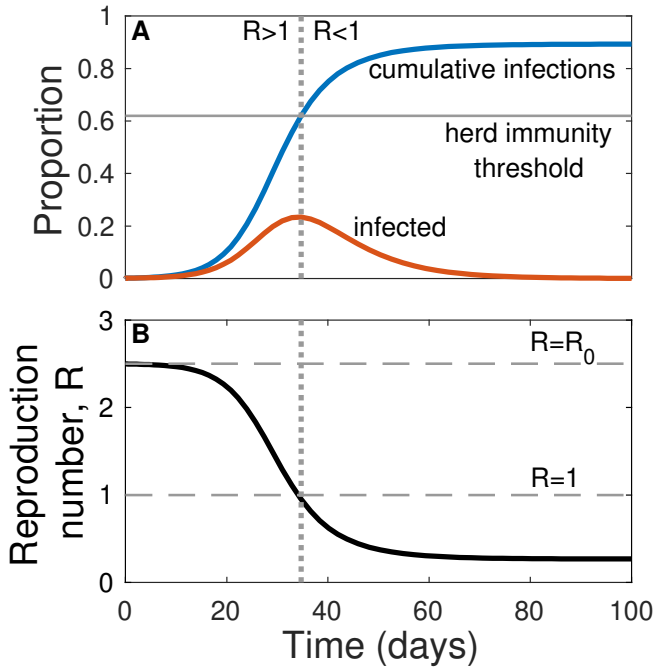

Figure 2. The reproduction number, *R*, and naturally acquired herd immunity.

The reproduction number, R, gives the average number of secondary infections produced by one infected individual. If $R > 1$ then the epidemic can grow and if $R < 1$ then the epidemic will shrink (A). As immunity accumulates in the population during the course of an epidemic, R declines from an initial value of R_0 (known as the 'basic' reproduction number), reaching $R = 1$ at the peak of the epidemic (B). In the absence of interventions (e.g. social distancing), naturally acquired herd immunity is therefore reached at the peak of the epidemic. However, individuals will continue to be infected as the epidemic declines (when $R < 1$), and so the final size of the epidemic may be much higher than the herd immunity threshold.

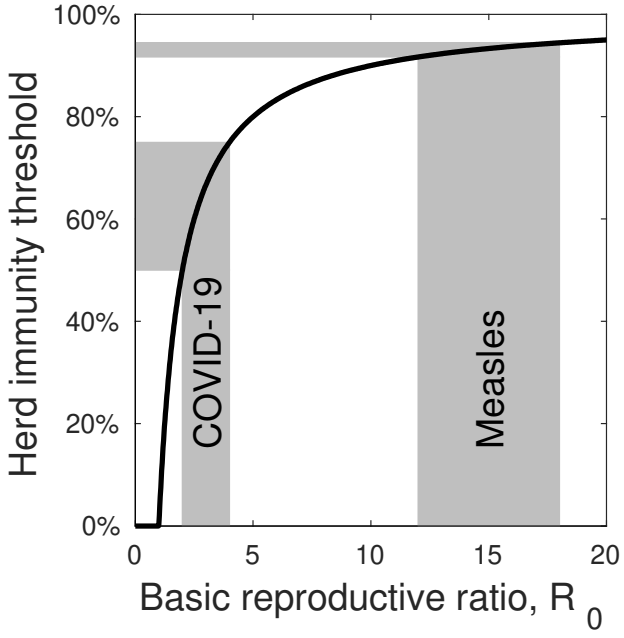

Figure 3. Relationship between the basic reproduction number, R_0 and the herd immunity threshold.

In a randomly mixing, homogeneous population, the herd immunity threshold is equal to $1 - 1/R_0$. The herd immunity threshold therefore initially increases rapidly for small values of R_0 but then slows down for larger values. Shaded regions illustrate estimated herd immunity thresholds for COVID-19 (with $2 < R_0 < 4$) and measles (with $12 < R_0 < 18$).

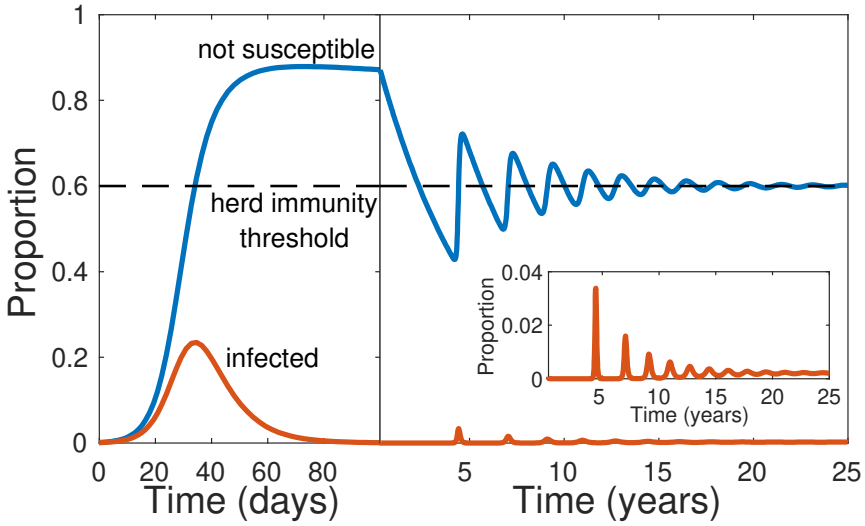

Figure 4. Loss of herd immunity can lead to subsequent epidemics.

Herd immunity status may be lost through population turnover, migration, waning immunity, and pathogen evolution. The population may eventually reach a stable equilibrium with the disease endemic.