Evolution, 2024, **78(6)**, 1092–1108
https://doi.org/10.1093/evolut/qpae041
Advance access publication 9 March 2024
Original Article

Endemic does not mean constant as SARS-CoV-2 continues to evolve

Sarah P. Otto[1], Ailene MacPherson[2], Caroline Colijn[2]

[1]Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[2]Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Corresponding author: Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. Email: otto@zoology.ubc.ca

Abstract

COVID-19 has become endemic, with dynamics that reflect the waning of immunity and re-exposure, by contrast to the epidemic phase driven by exposure in immunologically naïve populations. Endemic does not, however, mean constant. Further evolution of SARS-CoV-2, as well as changes in behavior and public health policy, continue to play a major role in the endemic load of disease and mortality. In this article, we analyze evolutionary models to explore the impact that a newly arising variant can have on the short-term and longer-term endemic load, characterizing how these impacts depend on the transmission and immunological properties of the variants. We describe how evolutionary changes in the virus will increase the endemic load most for a persistently immune-escape variant, by an intermediate amount for a more transmissible variant, and least for a transiently immune-escape variant. Balancing the tendency for evolution to favor variants that increase the endemic load, we explore the impact of vaccination strategies and non-pharmaceutical interventions that can counter these increases in the impact of disease. We end with some open questions about the future of COVID-19 as an endemic disease.

Keywords: evolution, epidemiology, model, SARS-CoV-2, selection

Introduction

Early in the global pandemic, COVID-19 levels rose and fell steeply, displaying rapid exponential growth and leading to widespread lockdowns and other public health measures to slow transmission (Ogden et al., 2022; Talic et al., 2021). The vaccination campaigns of 2021, followed by the nearly uncontrolled Omicron waves in early 2022 (Figure 1, BA.1 and BA.2 peaks), have now led to almost 100% immunological exposure in many countries. In Canada, for example, 100% of blood donors had developed antibodies to the spike protein from previous exposure to the virus by June 2023, with 80% also showing antibodies to nucleocapsid, indicating prior infection (Canadian Blood Services, 2023). The number of immunologically naïve individuals that fed COVID-19 dynamics throughout the pandemic has now greatly decreased, but in its place is a continual flow of newly susceptible individuals as humoral immunity wanes. For the past year, COVID-19 levels have ebbed and flowed in response to this waning and boosting of immunity, new variants, individual behavior, and changing public health measures. These peaks and troughs are more subdued wavelets, compared to earlier Omicron peaks (Figure 1).

COVID-19 is now considered an endemic disease, being both widespread and persistent, adding to the respiratory infectious diseases with which we must routinely contend. Its now-endemic nature reflects a balance between waning immunity and ongoing transmission, leading to a turnover of cases across the globe. Endemic does not mean "constant," as new variants and behavioral shifts drive change. Endemic also does not mean "rare," as waning and transmission rates have remained high, and waves continue to be driven by new variants (e.g., Figure 1). Here we explore mathematical models to improve the understanding of how the ongoing evolution of SARS-CoV-2, as well as our behavioral responses, will shape endemic COVID-19 and similar diseases.

When most individuals in a population are susceptible (epidemic phase), any variant or behavioral measure that affects the transmission rate will have a direct effect on the number of new infections over the short term, as exposures determine the spread of disease. When a disease first appears, the reproductive number describing the number of new infections per infection, R_0, is given by the transmission rate divided by the clearance rate of the infection in the classic SIR epidemiological model (Keeling & Rohani, 2011). Variants that increase transmission or behavioral changes that reduce transmission directly impact the number of new infections and the rate of exponential growth, but these new infections have little immediate effect on the large pool of susceptible individuals. Evolutionary models of this epidemic phase (e.g., Day et al., 2020) thus typically focus on capturing the complexities of transmission, for example, including exposed or asymptomatic classes and non-homogenous mixing, but often ignore waning of immunity or the possibility that variants may evade any such immunity earlier.

By contrast, when a disease is endemic and most individuals have some degree of immunity, waning must be explicitly considered and modeled in a manner that allows variants to infect earlier (as in the SIR_n model, with multiple recovered

Received September 28, 2023; revisions received February 24, 2024; accepted March 7, 2024
Associate Editor: Tim Connallon; Handling Editor: Jason Wolf

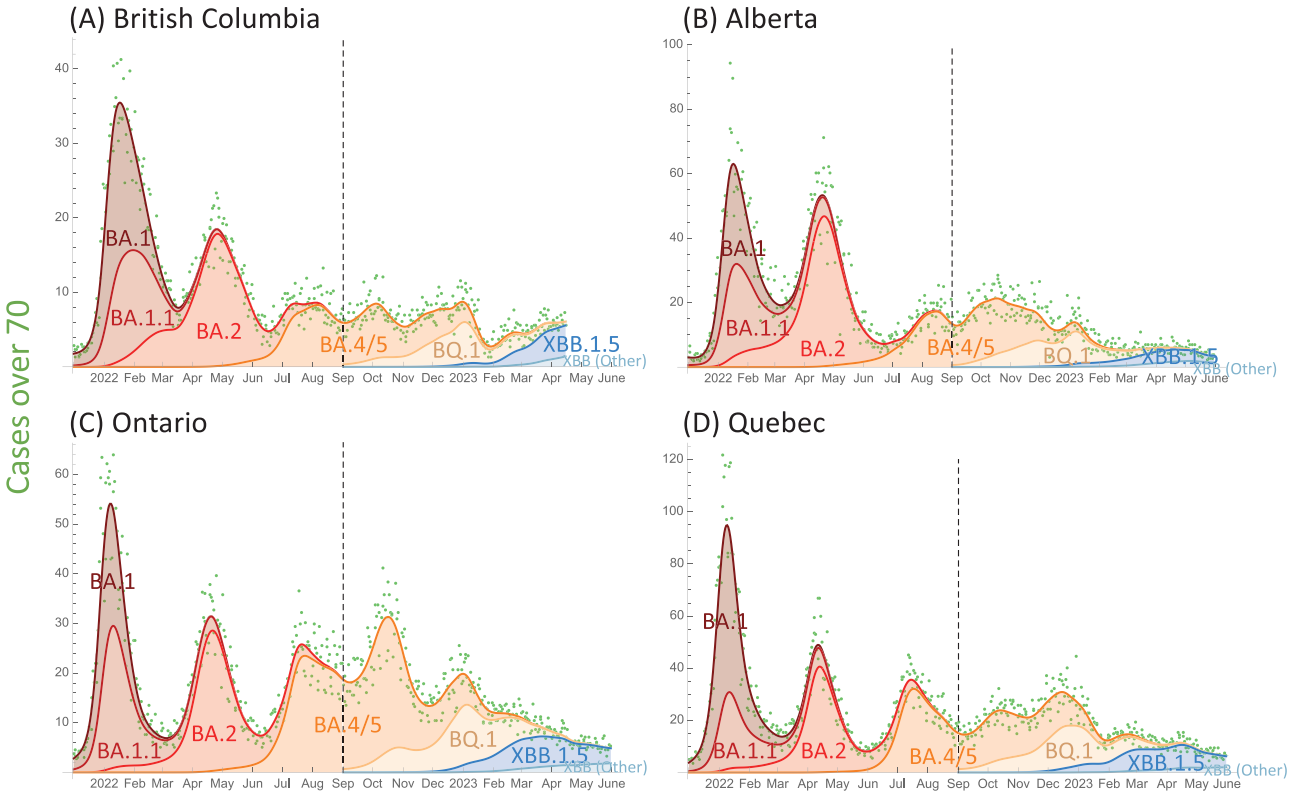

Figure 1. COVID-19 trends across four provinces in Canada. Major waves in early 2022 were driven by the rise and spread of Omicron, whose immune-evasive properties allowed widespread infection at a time when public health measures were largely relaxed (peak in January 2022: BA.1, April: BA.2, and July: BA.4 and BA.5). A year later, Omicron variants have continued to spread rapidly (peak in December 2022: BQ.1; April 2023: XBB.1.5), but they no longer cause major waves in cases. PCR-confirmed cases per 100,000 individuals aged 70 + (dots) are used to illustrate case trends, as testing practices changed dramatically over this time period but this age group remained eligible for testing. To guide the eye, a cubic spline fit (lambda = 3) was applied (top curves in each panel), and the frequency changes of each variant under this curve were fitted by maximum likelihood using duotang (CoVaRR-Net's CAMEO, 2023). Genomic sequence data from each province were obtained from the Canadian VirusSeq Portal (VirusSeq, 2023) and fit by maximum likelihood to a model of selection in two periods: first 9 months using BA.1 as a reference (left of dashed line); second 9 months using BA.4/5 as a reference (right of dashed line), grouping all clades within a family together except when a subclade is also mentioned (e.g., BQ.1 separated from BA.5). See Supplementary *Mathematica* package for scripts and duotang (CoVaRR-Net's CAMEO, 2023) for methodological details and finer resolution of lineages and time periods.

classes, that we consider here). Within this endemic context, many individuals have had prior exposure to the disease, and only some are currently susceptible due to waning immunity. Variants that escape immunity can infect earlier in this waning process, gaining a selective advantage that is absent during the initial epidemic phase when few individuals are immune. Evolutionary or behavioral changes in transmission rate are also dampened during the endemic phase because faster transmission depletes the limited pool of susceptible individuals. Here, we focus on evolution during the endemic phase by using a model intended to capture the nuances of immune waning and reexposure through infection or vaccination.

At one extreme, infections during the endemic phase may be limited not by transmission but by waning, which has been used as an argument against non-pharmaceutical measures like masking by some public health officials, e.g.:

> Masks "can delay transmission, they can reduce transmission, but they're not actually effective measures at a population level," because "exposure is essentially universal now to COVID-19."
>
> November 2, 2022 in Today in BC podcast

This argument assumes that transmission is so common that reducing risks of exposure no longer matters, as another exposure will occur soon thereafter. This is a strong claim, with major implications for both individual and public health decisions. It is essentially a claim that endemic levels of a disease such as COVID-19 are not under our control.

Mathematical models can help evaluate such claims and determine whether and to what extent our actions affect the endemic load of disease and mortality. Models can also predict how this load would change in the face of new variants and how this depends on the properties of those variants. Models can also help evaluate plausible parameter ranges, allowing us to assess how much reducing transmission would affect population-wide incidence. Here, we tailor standard epidemiological models to the current phase of COVID-19 to better understand the risks posed by new variants and our ability to control endemic diseases.

Model background

We use a classic compartment model, SIR_n, as illustrated in Figure 2, measuring the fraction of the population that is

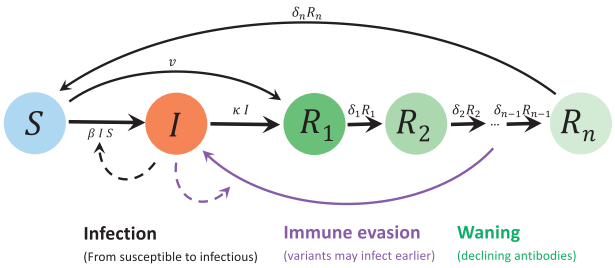

Figure 2. Epidemiological model used to predict impact of changing variants, behavior, and policy on endemic levels of disease. We consider populations that have a high level of immunity due to prior infection and/or vaccination and that consist of S: a susceptible fraction, I: an infected fraction, and R_j: a recovered fraction with immunity at different stages of waning. Parameters are β: transmission rate, κ: recovery rate, δ_j: per-class waning rate per day, and ν: vaccination rate at the population level, all measured in the present-day population with prior exposure. Movement between adjacent recovered classes is set equal to $\delta_j = n\delta$, so that the expected time between first recovering and returning to the susceptible state is $1/\delta$ days. Coinfection and evolutionary dynamics within hosts are ignored.

in the susceptible class (S), the infectious class (I), or in one of several recovered classes (R_j). These sequential recovered classes allow for different stages of waning immunity (i, ranging from 1 to n) and capture the dynamics of neutralizing antibodies that help protect against infection (Andrews et al., 2022; Khoury et al., 2021). When measured on a log scale, neutralizing antibodies rise to a high level soon after infection or vaccination and then decline linearly over time since vaccination and/or infection (e.g., Lau et al., 2021; Lau et al., 2022; Evans et al., 2022; Jacobsen et al., 2023). We thus consider the R_j classes to fall along different stages of this decline (R_1 being highest and R_n lowest), roughly representing classes of individuals with similar log antibody concentrations. As antibody levels fall over time, individuals move from one class to the next (R_j to R_{j+1}) until levels are so low that infection is no longer prevented (R_n waning to S).

As we are modeling the long-term epidemiological dynamics in a population previously exposed to the virus via vaccinations and/or infections, we emphasize that the susceptible class, S, consists of individuals who have had previous exposure but are currently susceptible due to waning immunity. Throughout this article, we are thus describing the epidemiological dynamics in a previously challenged population. We assume that births and deaths involve a small subset of the population and do not include their influence on the disease dynamics. Using this waning immunity model, we examine the effect of variants, vaccines, and non-pharmaceutical interventions (NPIs), on endemic disease dynamics, as captured by the equilibrium prevalence and transient cycles. While the quantitative outcome of these dynamics may be impacted by population heterogeneity and assortative mixing, we focus primarily on the qualitative patterns caused by variants and public health interventions, focusing on heterogeneity in immune status.

When we model vaccination, vaccines move individuals from the susceptible (S) to the first recovered class (R_1) at a constant rate ν (Figure 2). We consider a constant vaccination rate rather than a per-capita rate to describe vaccine programs that are either supply-limited or that aim to meet a target vaccination rate within a jurisdiction. We also assume that vaccinations are targeted primarily to susceptible individuals, given widespread recommendations that vaccines are unnecessary for those with recent infections. Because low testing rates make it challenging to determine an individual's disease history, we also consider the case where vaccinations are given regardless of an individual's immunity status. Vaccination is assumed to be protective against infection in this model until vaccine-induced immunity wanes, which occurs at the same rate that infection-induced immunity wanes (although we do extend the model to consider the possibility that vaccination does not elicit an immune reaction in some individuals). We do not separately model disease or severity, or vaccine's effectiveness against these as distinct from protection against infection, although the dynamics of severe cases are expected to mirror caseloads for variants of similar severity.

Dynamics

Before considering variants or NPI measures, the dynamics describing changes in the number of individuals in each compartment within the SIR$_n$ model are (see Figure 2 for a description of the parameters):

$$\begin{aligned}\frac{dS}{dt} &= \delta_n R_n - \beta SI - \nu \\ \frac{dI}{dt} &= \beta SI - \kappa I \\ \frac{dR_1}{dt} &= \kappa I + \nu - \delta_1 R_1 \\ \frac{dR_j}{dt} &= \delta_{j-1} R_{j-1} - \delta_j R_j \qquad \text{for } 2 \le j \le n.\end{aligned} \tag{1}$$

These dynamics approach an equilibrium, which can be determined by setting the derivatives to zero and solving for the fraction of individuals in each class. This results in two equilibria; one corresponds to the disease being absent ($\hat{S} = 1$) and the other to the disease being endemic:

$$\begin{aligned}\hat{S} &= \frac{\kappa}{\beta} \\ \hat{I} &= \left(1 - \frac{\kappa}{\beta}\right)\frac{\delta}{\delta + \kappa} - \frac{\nu}{\delta + \kappa} \\ \hat{R}_j &= \frac{1}{n}\left(\left(1 - \frac{\kappa}{\beta}\right)\frac{\kappa}{\delta + \kappa} + \frac{\nu}{\delta + \kappa}\right) \qquad \text{for } 1 \le j \le n.\end{aligned} \tag{2}$$

Importantly, because we are explicitly modeling endemic COVID-19, most individuals have previously been exposed to SARS-CoV-2, susceptibility and infectiousness may be lower in the current population than when the virus first appeared in humans because of cellular immunity, any residual humoral immunity among susceptible individuals (Tan et al., 2023), and/or due to any behavioral changes (including better ventilation, testing, and self-isolation practices). For example, the rapid induction of cellular immunity reduces the viral load of typical breakthrough infections (Puhach et al., 2022), lowering transmission (β) compared to a fully naïve population. The epidemiological dynamics in this endemic model thus depend on an "*endemic basic reproductive number*," $\tilde{R}_0 \equiv \beta/\kappa$, which is the basic reproductive number in a population consisting entirely of currently susceptible, but previously vaccinated or infected, individuals whose immunity has waned. In contrast to the initial R_0 for COVID-19

at the time of its emergence (median R_0 estimated at 3.6–6.1 in different European countries and the United States (Ke et al., 2021); ~3 in British Columbia Canada (Anderson et al., 2020); 1.5–5 in analyses of Iran, China, Italy, and South Korea (Aghaali et al., 2020)), the parameters of the endemic model and in $\tilde{R}_0$ (transmission, β, and recovery, κ) refer to rates in this previously challenged population. Our estimates of $\tilde{R}_0$ range from 1 to 6 (Appendix A), depending on estimates used for recovery rates, waning, and the endemic level of infections within a population, with a value of $\tilde{R}_0 \approx 2$ for the parameters considered typical (Supplementary Table S1).

If this previously challenged population were fully susceptible ($\hat{S}$ near one), the disease would spread when rare as long as transmission rates were higher than recovery rates, $\tilde{R}_0 = \frac{\beta}{\kappa} > 1$, which we assume to hold. In this case, the endemic equilibrium in *Equation 2* exists and is stable for all examples considered here. The endemic equilibrium may, however, be unstable (Hethcote, Stech & Van Den Driessche, 1981), leading to sustained cyclic dynamics, outside of the parameters used here (e.g., for n large enough).

The equilibrium can also be written in terms of $\tilde{R}_0$ as:

$$\hat{S} = \frac{1}{\tilde{R}_0}$$

$$\hat{I} = \left(1 - \frac{1}{\tilde{R}_0}\right)\frac{\delta}{\delta+\kappa} - \frac{\nu}{\delta+\kappa} \qquad (3)$$

$$\hat{R}_j = \frac{1}{n}\left(\left(1 - \frac{1}{\tilde{R}_0}\right)\frac{\kappa}{\delta+\kappa} + \frac{\nu}{\delta+\kappa}\right) \qquad \text{for } 1 \le j \le n.$$

Also of relevance is the number of bouts of disease that an individual expects per year, which is $365\ \beta\ \hat{S}\ \hat{I} = 365\ \kappa\ \hat{I}$, assuming average behavior (see Supplementary Table S1).

Given that recovery rates are higher than waning rates ($\delta \ll \kappa$), *Equation 3* shows that the number of infectious individuals at the endemic equilibrium is reduced by the number of vaccinations within a typical recovery period (ν/κ). If more vaccinations were to be given in a typical recovery period than the fraction of individuals expected to be infectious in the absence of vaccination, the disease could be driven extinct locally (though we note that in this model vaccination has a very high, if temporary, efficacy against infection, and that reintroductions are expected from importations, animal reservoirs, and chronic infections). Uptake of additional vaccine doses during 2023 has, however, been so low in many countries as to make little difference to the incidence and dynamics of SARS-CoV-2 (e.g., daily [annual] rates of 0.012% [4.7%] in France and 0.023% [8.8%] in the United States from January 1 to April 30, 2023; Our World in Data, 2023). To simplify the discussion, we ignore ongoing vaccination for now, returning later to a discussion of the impact that vaccination uptake can have on individual risks of infection and on the overall incidence of disease.

Spread of a variant during the endemic phase

A new variant may spread within the population if it is more transmissible (e.g., better binding to ACE2 receptors on host cells), more immune evasive, or both (see Cao et al., 2023 for empirical measures for SARS-CoV-2). We can calculate the rate of spread of a variant using the SIR_n model by allowing different transmission rates for the resident variant (β) and the new variant ($\beta^* = \beta + \Delta\beta$) and by allowing immune evasive variants to infect earlier than the resident strain, while antibody levels are at intermediate levels. Specifically, we assume that an immune evasive variant can infect the last m recovered classes (each of which is at frequency $\hat{R}_j$ at the endemic equilibrium given by *Equation 3*), as well as susceptible individuals.

Short-term impact of a variant

Using local stability analysis, we explore the short-term spread of a variant that arises during the endemic phase, assuming that the fractions in each class are near the equilibrium given by *Equation 2*, and then we determine the long-term impact on infection rates following the spread of the variant. We expect the results assuming an equilibrium to be robust to the churning spread of variant-after-variant (Figure 1) as long as recent waves have not led to major departures from the fractions of individuals in each class at the endemic equilibrium (Figure 2). We investigate the robustness of our results to this and other assumptions in Appendix B.

As described in Appendix A, a new variant introduced into a population near the endemic equilibrium has a selective advantage of:

$$s = \underbrace{\frac{\Delta\beta}{\beta}\,\kappa}_{\text{Transmission advantage}} + \underbrace{m\,\hat{R}_n\,\beta^*}_{\text{Evasion advantage}} \qquad (4)$$

The selection coefficient, s, describes the rate at which the new variant spreads relative to the resident variant. Selection coefficients describing evolutionary changes in SARS-CoV-2 have been estimated in many jurisdictions using sequence information and are often relatively stable over time and space when measured consistently against the same reference strain (van Dorp et al., 2021; Otto et al., 2021).

Long-term impact of a variant

What are the consequences of a spreading variant for the incidence of disease? The incidence is initially expected to rise exponentially at a rate proportional to selection (specifically, $s\ \hat{I}$), but this rise in cases is only temporary as the new variant spreads through the currently susceptible population. Over the long term, we show that the impact on the endemic level of disease depends strongly on whether the variant increases transmission rates and/or increases immune evasiveness, even for variants with the same selective advantage. Among immune evasive variants, the impact on the long-term caseload depends strongly on the transience or persistence of immune evasion (from the perspective of the virus) during subsequent infections. Transient evasiveness is expected if, during infection, new variant-specific antibodies are elicited that fully recognize the new variant, whereas persistent evasiveness is expected when there is strong immune imprinting, such that infection with the variant leads primarily to the proliferation of pre-existing antibodies ("back-boosting") rather than an expansion of the memory B cell repertoire to specifically target the new variant (see Zhou et al., 2023 for a review of evidence on immune imprinting with SARS-CoV-2). Metanalyses, for example, find that Omicron infections are less protective against reinfection with Omicron compared to the protective effects of pre-Omicron infections, suggesting some persistence of Omicron's immune-evasive properties (Arabi et al., 2023).

In particular, when the population is comprised entirely of the new variant, the altered transmission rates and immune evasiveness cause the endemic level of disease to change from the original model to:

$$\hat{I}^* = \left(1 - \frac{\kappa}{\beta + \Delta\beta}\right) \frac{\delta + \Delta\delta}{\delta + \Delta\delta + \kappa} \tag{5}$$

(found by solving *Equation A1* for the endemic equilibrium when only the variant is present). The term $\Delta\delta$ refers to how the variant changes the rate of complete waning, from first entering the recovered class to returning to the susceptible state (i.e., $1/(\delta + \Delta\delta)$ is the mean number of days to return to susceptibility). (As short-hand, we refer to a variant's impact on immune evasion as a change in the waning rate $\Delta\delta$, but the model actually assumes log-antibody levels wane at a constant rate but the variants can just infect earlier, becoming susceptible sooner.)s

Equation 5 allows us to evaluate the long-term impact of different types of variants. For an immune evasive variant, the results are strongly dependent on the variant-specific immunity that develops after infection, even if the lineages have the same selective advantage and rate of spread (s, *Equation 4*), and so cause the same initial rise in cases. Consider two extreme possibilities:

- **Transient immune evasiveness:** If a variant better evades the initial suite of antibodies but causes infections that generate variant-specific immunity, subsequent infections may no longer be immune evasive. In this case, subsequent infections would require the full waning period (returning to the S compartment and not the R_i compartments), as for the resident strain. With only transient evasiveness, the long-term level of COVID-19 is ***unaffected*** by a new variant ($\Delta\delta = 0$).
- **Persistent immune evasiveness:** If a variant allows infections to occur earlier, both in the first and in subsequent infections, then the rate of return to susceptibility is consistently higher ($\Delta\delta = \frac{m}{n-m}\delta$). With waning slow relative to recovery ($\delta + \Delta\delta \ll \kappa$), a persistently immune evasive variant causes the incidence of disease to ***rise in proportion*** to the increased rate of waning, $\hat{I}^* \approx \hat{I}\ (1 + \Delta\delta/\delta)$.

Figure 3 illustrates the temporal dynamics of a variant initially at 1% frequency, obtained by numerically integrating the differential equations in system (*Equation 1*). In either case, an immune-evasive variant spreads in the short term because of the selective advantage, s, gained by infecting susceptible individuals earlier, causing an immediate rise in cases, but the subsequent dynamics and long-term impact differ greatly, depending on whether variant-specific immunity builds. The extent to which exposure to a variant elicits variant-specific humoral or cellular immunity almost certainly falls between these two extremes.

By contrast, a more transmissible variant changes the endemic equilibrium to:

$$\hat{I}^* = \hat{I}\ \left(1 + \frac{1}{\tilde{R}_0 - 1}\ \frac{\Delta\beta}{\beta + \Delta\beta}\ \right) \tag{6}$$

Summarizing our results for a more transmissible variant:

- **Higher transmissibility:** If a variant increases transmission rate, spread in the short term depends directly on the change in transmission ($s = (\Delta\beta/\beta)\ \kappa$; *Equation 4*), while the long-term impact on the incidence of disease exhibits diminishing returns ($\hat{I}^*/\hat{I}$ depends on $\Delta\beta/(\beta + \Delta\beta)$). Thus, a more transmissible variant has **a less than proportionate influence** on the number of cases in the long term, unless the susceptible pool is large and $\tilde{R}_0$ small ($\tilde{S} = 1/\tilde{R}_0 > 1/2$).

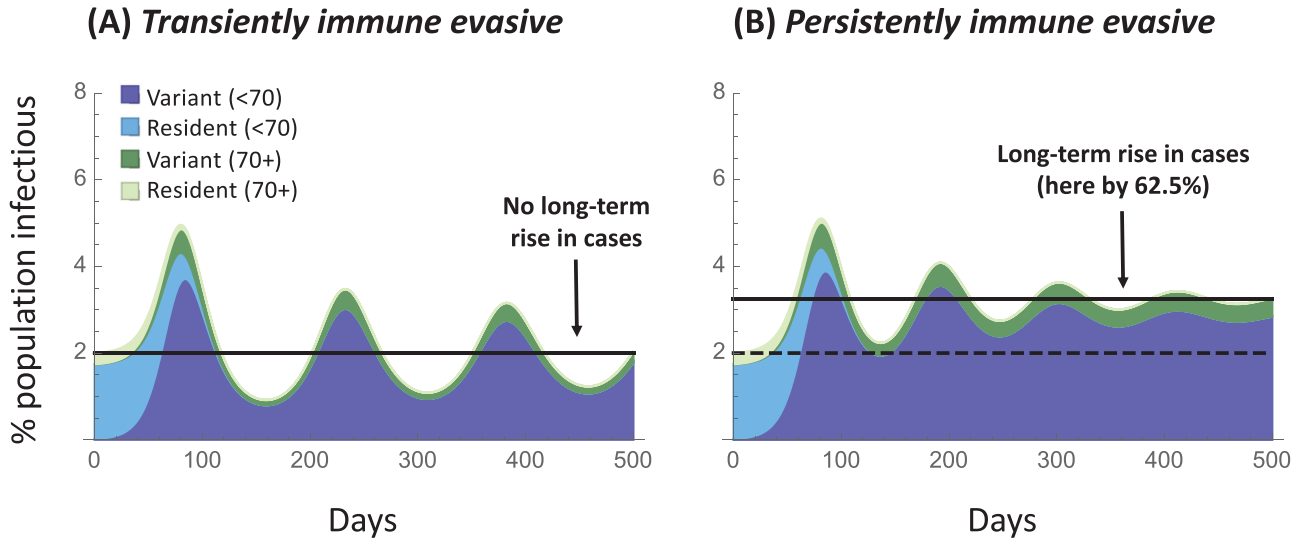

Figure 3. Impact of the spread of an immune-evasive variant depends on whether a variant-specific immune response is elicited. Plots illustrate the dynamics over time of a more immune evasive variant, which is able to infect earlier in the waning period (by $m = 2$ out of $n = 5$ recovered classes), giving the variant an $s = 8.3\%$ selective advantage per day, which lies in the range of the faster spreading variants observed in the past year (CoVaRR-Net's CAMEO, 2023). While the short-term spread of the variant (dark shading taking over from light shading) and rise in cases are nearly identical (given s is the same), the long-term consequences differ substantially depending on whether the variant's evasive properties are (panel A) transient or (panel B) persistent (dashed and solid lines represent the equilibrium before and after the variant spreads). The endemic equilibrium rises only if evasiveness persists in subsequent infections (panel B). We illustrate the dynamics in younger (under 70) and older (70+) individuals, who are more prone to severe cases. Parameters: $\kappa = 0.2$, $\delta = 0.008$, $\hat{I} = 2\%$, $\beta = 0.42$, the nominal parameter estimates given in Appendix A for all age classes.

While a more immune evasive variant increases the pool of susceptible individuals available to them (in our model, by adding the last m recovered classes to the susceptible class), a more transmissible variant depletes the susceptible pool. This can be seen by the effect on the susceptible class at equilibrium, which decreases from $\hat{S} = \kappa/\beta$ for the resident virus to $\hat{S}^* = \kappa/(\beta + \Delta\beta)$ for a more transmissible virus. For this reason, a more transmissible virus is, to some extent, self-limiting and often has less of an effect on the long-term number of cases than seen for a permanently immune evasive variant. This is illustrated in Figure 4, which shows that the equilibrium rises less for a given % increase in transmissibility (panel C) than for the same % increase in waning rate for a permanently immune evasive variant (panel B), unless $\tilde{R}_0$ is so small that most individuals in the population are susceptible. Of course, variants may combine features affecting transmissibility and immune evasiveness (Cao et al., 2023), as explored in Supplementary Figure S1.

This model illustrates that the spread of a variant can contribute to epidemic waves, like those seen over the past year in Figure 1, in two ways, either through the initial replacement of the resident strain (the first cycle shown in Figures 3 and 4) or through the continued ebb and flow of incidence due to immune waning ("echo" waves in later cycles). The frequency at which new beneficial variants arise determines the relative contribution of each of these. Due to the substantial population-level immunity that persists in the population at the endemic equilibrium, the model predicts that these waves are modest, at most doubling incidence, a pattern consistent with the wavelets observed since the summer of 2022 (Figure 1) and in contrast to the dramatic waves of earlier variants during the epidemic phase when immunity was limited. The lower magnitude of peaks is expected at the endemic equilibrium because the effective reproductive number for the resident must be one ($R_{e,t} = 1$) compared to $R_{e,t} = 1 + \Delta\beta/\beta$ for a more transmissible variant (e.g., $R_{e,t} = 1.42$ in Figure 4). Thus, only a modest drop in the susceptible population is needed ($1/R_{e,t}$) before the infectious class peaks and falls again.

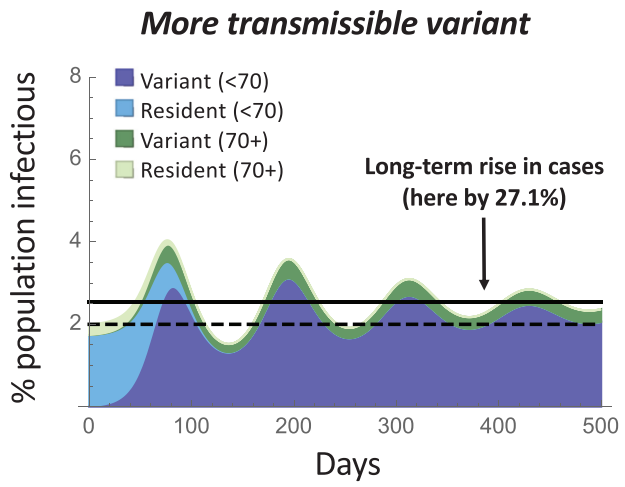

Figure 4. Impact of the spread of a more transmissible variant. Plot illustrates the dynamics over time of a more transmissible variant, which increases β (and hence $\tilde{R}_0$) by 42%, chosen to give the variant the same selective advantage as in Figure 3 (s = 8.3%). While exhibiting a similar short-term rise in cases as in Figure 3, the long-term impact is intermediate. Parameters are identical to Figure 3, with $\beta^* = 0.59$.

Robustness of results

The robustness of these results is explored in Appendix B, considering different models of immunity (including leaky immunity) and the inclusion of features such as an exposed class and failure to seroconvert. The strength of selection (*Equation 4*) and the long-term impact on endemic levels of disease (*Equation 5*) are robustly observed. The speed at which the waves dissipate over time, however, is sensitive to model assumptions, stabilizing faster than observed above in many cases (Supplementary Figure S2), so we caution that the nature of subsequent oscillations ("echo" waves) caused by the spread of a variant is hard to predict.

Up until now, we have focused on a single variant to isolate its impact. In Supplementary Figure S4, we explore the spread of a secondary variant soon after a previous variant has spread to determine the robustness of the results to violations of the assumption that the system starts at equilibrium. Considering a more transmissible variant followed by a persistently invasive variant (top row) or vice versa (bottom row), we find that the initial spread and the long-term dynamics continue to be well predicted by *Equations 4 and 2*, respectively, although selection varies slightly depending on the phase of oscillations caused by the previous variant.

We conclude that variants may have dramatically different long-term impacts on the level of disease depending on the nature of the advantage (transiently or persistently immune evasive and/or more transmissible), despite exhibiting the same selective advantage and hence spreading at the same rate (e.g., with selection of s = 8.3% per day in Figures 3 and 4). Indeed, a variant that is transiently immune evasive but less transmissible can spread and would be expected to reduce the equilibrium level of disease, except that once immunity to this variant has built, the previous resident reemerges because of its higher transmissibility (Supplementary Figure S1B).

Figure 5 shows these long-term impacts on disease incidence at the endemic equilibrium across the range of plausible parameters (Appendix A). A transiently immune evasive variant has no long-term impact (panel A), whereas the rise in cases is nearly proportional to the ability of a variant to evade immunity, if that evasiveness is persistent, regardless of the exact parameter values (panel B). By contrast, the long-term impact of a more transmissible variant depends strongly on the current transmissibility, as measured by the endemic reproductive number. The larger $\tilde{R}_0$ is, the smaller the long-term impact of a more transmissible variant is on disease levels (Figure 5C), essentially because the pool of susceptible individuals is then smaller and rapidly depleted by a more transmissible variant ($\hat{S} = 1/\tilde{R}_0$). That said, for given waning (δ) and recovery (κ) rates, the endemic level of disease is higher when $\tilde{R}_0$ is higher (*Equation 3*), so a small percentage increase in disease incidence can still have a numerically important impact on the burden of disease.

Measures to counteract rises in disease incidence

In the face of variants that are increasingly transmissible and/or persistently immune evasive, the endemic level of disease is expected to rise over time, but these increases can be countered by protective measures at the individual and population levels. Protective measures range from vaccination to NPI measures, such as testing and self-isolation, avoiding crowded indoor spaces, improving ventilation, and the wearing of well-fitting

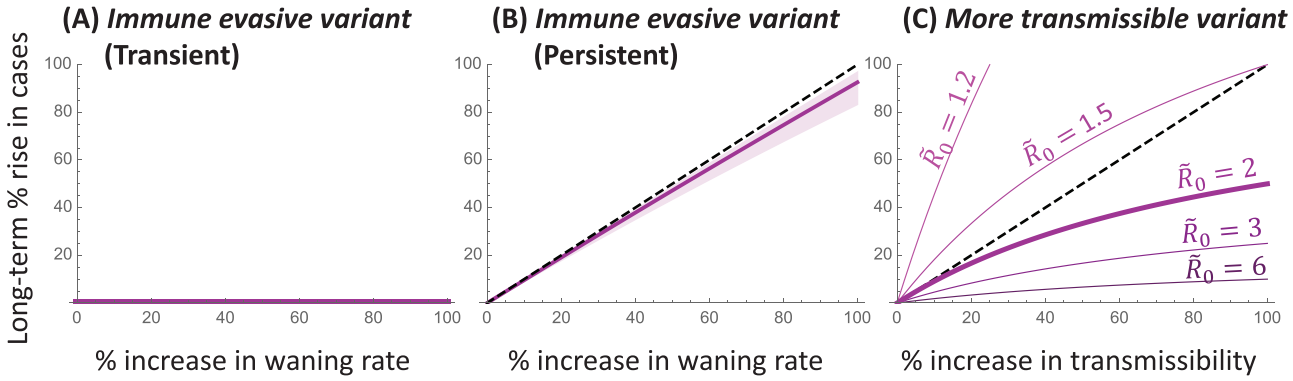

Figure 5. Long-term impact of a variant. The percent change in the endemic equilibrium is shown as a function of the percent by which the variant increases the rate at which recovered individuals become susceptible again ($\Delta\delta$, panels A and B) or transmissibility ($\Delta\beta$, panel C). A transiently immune-evasive variant has no long-term impact, while a persistently immune-evasive variant causes the endemic incidence of disease to rise nearly in proportion (dashed curve) across all parameters considered plausible (shading, with the nominal parameter values illustrated by a thicker curve; see Appendix A and Supplementary Table S1). By contrast, the impact of a more transmissible variant that increases β depends strongly on $\tilde{R}_0$ (but none of the other parameters), leading to a less than proportional rise in cases whenever $\tilde{R}_0 \geq 2$ (see *Equation 6*).

and high-quality masks (The Royal Society, 2023). Here we explore the impact of these protective measures at both the individual level, modulating the frequency of infections, and the population level, modulating the endemic incidence of disease. Many previous models have explored vaccination and NPIs (e.g., Scherer & McLean, 2002; Eikenberry et al., 2020; Mulberry et al., 2021); our purpose here is to explore how the effects of variants can be counteracted by our actions, using the same models and parameter ranges explored above (Appendix A).

Vaccination

Vaccination allows individuals to short-circuit the disease cycle by boosting antibody levels and immunity by another dose rather than by infection. Globally, 65% of people have had the primary series of COVID-19 vaccines but an average of only 0.35 booster doses have been distributed per person (Our World in Data, 2023; accessed August 22, 2023). Jurisdictions vary widely in recommended vaccine schedules and access to vaccines. For example, only individuals at higher risk of serious illness are eligible for COVID-19 boosters in the United Kingdom (NHS, 2023). In Canada, the National Advisory Committee on Immunization recommends that all adults be offered vaccines 6 months after the last dose or infection (NACI, 2023). The Centre for Disease Control in the USA, however, recommends that individuals stay up to date with important vaccine updates (e.g., the updated mRNA vaccines providing protection against BA.4 and BA.5 in the fall of 2022 and against XBB in the fall of 2023; CDC, 2023).

There is substantial uncertainty and confusion in both public and public health circles about the value of regular vaccinations against COVID-19 (Lin et al., 2023). Here, we explore one aspect: how much do regular vaccinations reduce the burden of disease expected near an endemic equilibrium?

We consider the impact of policies aimed at future vaccine uptake, encouraging vaccination of a portion of the population (ν) per day. Given that vaccines are recommended only after a substantial amount of time has passed since the previous dose or infection, these vaccines are assumed to target individuals in the susceptible class, moving them into the first recovered class (S to R_1; *Equation 1*). Unlike many previous epidemiological models (reviewed in Scherer & McLean, 2002), we assume that the target vaccination rate is set by policy, adjusting public health campaigns, vaccine cost, and availability to meet these targets (i.e., dS/dt in *Equation 1* declines by a fixed daily rate, ν, rather than a per capita rate, $\nu\ S$). This model choice also allows us to model the distribution of a specific number of vaccines, whereas vaccinating susceptible individuals at a per capita rate causes vaccine coverage to vary over time with S (see Appendix B for alternative models).

We consider vaccination rates in Canada as typical of what can be achieved when vaccines are available at regular intervals (every 6 months). From April to July 2023, vaccination rates in Canada averaged only 0.012% of the population per day (annual rate of 4.5%; Health Infobase Canada, 2023). While these vaccinations help those individuals receiving a dose, this level has a negligible impact on the endemic level of cases (decreasing $\hat{I}$ from 2% to 1.94% for the nominal parameter values). Many public health agencies have encouraged COVID-19 vaccine updates in the fall (Mahase, 2023). For example, vaccination rates in Canada during September–December 2022 were 14 times higher (0.174% of the population per day, an annual rate of 63.5%; Health Infobase Canada, 2023), a rate that substantially lowers the endemic equilibrium level if maintained (from 2% to 1.16% for the nominal parameter values).

At an individual level, vaccination reduces the number of infections that one expects to have. Individuals on a regular 6-month vaccination schedule are expected to be protected from neutralizing antibodies for $1/\delta$ out of every 180 days. Calculating the probability of waning and infection before their next vaccine (*Equation A2*; Appendix A), a regularly vaccinated individual expects to have about 60% as many infections per year (0.88 vs 1.46) for the nominal parameter values (Appendix A). Across the range of parameters considered plausible, vaccination every 6 months leads to only 40%–66% as many infections annually (Supplementary *Mathematica* package).

At a population level, in our model, vaccination reduces the endemic level of infections to $\hat{I}_\nu = \hat{I} - \frac{\nu}{\delta+\kappa}$ (*Equation 2*). That is, the endemic level of infections is reduced by approximately the number of vaccinations conducted during the infectious period (ν/κ, given that waning is considerably slower than recovery). Figure 6 illustrates the impact of increasing and

maintaining the vaccination rate at the higher levels observed in Canada in the fall of 2022 ($\nu = 0.174\%$). In this case, the long-term incidence of infection can be driven down substantially (panel A). Across the range of parameters considered in Appendix A, this population-level benefit leads to a 12% to 100% decline in incidence of disease, falling at the lower end of the benefit when disease incidence is high without vaccination ($\hat{I} = 4\%$) but at the higher end and allowing complete eradication when disease incidence is low without vaccination ($\hat{I} = 0.5\%$).

One policy option considered in many jurisdictions is to regularly vaccinate only the more vulnerable segment of the population. Figure 6B illustrates, however, that limiting vaccination to the more vulnerable population (shown here as vaccinating only those over 70 at a rate $\nu = 0.174\%$) has less impact on the frequency of infections experienced by this vulnerable population, because the incidence of COVID-19 remains high overall, increasing their risk of exposure.

That said, the additional protection provided by COVID-19 vaccines against severe disease, above and beyond the protection provided against infection, means that the risk of hospitalization and death can be lowered by vaccinating the vulnerable (Chemaitelly et al., 2022; Nyberg et al., 2022). Further reducing the risk of infection and severe disease, however, requires a broader vaccination campaign. Broad vaccination campaigns provide additional protection for the vulnerable, while also reducing the number of sick days, risks of long COVID, and severe disease among those not known to be vulnerable.

As noted in Appendix B, seroconversion rates upon vaccination are high (Wei et al., 2021), so we do not correct ν for the small fraction of doses that fail to elicit an immune response. Not all individuals will, however, achieve high levels of immunity following vaccination and not all will be protected from infection. A mixture of induced immunity could be modeled by moving vaccinated individuals into a distribution of R_j classes. In addition, some individuals being vaccinated may have been exposed in the recent past (i.e., coming from the infectious or recovered compartments, not solely from the susceptible classes). These possibilities are expected to lower the protection offered by vaccination, because a fraction of vaccines are distributed to individuals who are already immune and so would require higher vaccination uptake to achieve the benefits described above (see Appendix B for an example where vaccinations are distributed regardless of disease status).

NPI measures

A wide variety of non-pharmaceutical interventions have been deployed to counter the spread of SARS-CoV-2, including testing and self-isolation, enhancing ventilation and air filtration, and wearing of high-quality masks (see evaluation of evidence in the report by The Royal Society, 2023). Here, we consider the individual-level and population-level benefits of NPIs, as a function of their impact on preventing transmission of the virus, modeled by NPIs preventing a portion p of transmissions both from and to NPI users. Specifically, we assume that the NPI measures reduce transmission from β to $(1-p)\ \beta$ if one member in an interaction practices the measures and to $(1-p)^2\ \beta$ if both do. Recent meta-analyses suggest that properly wearing high-quality masks, for example, protects against infection by $p = 25\%$–50% (Appendix C). For the sake of comparison to the impacts of variants, we explore a limited model, ignoring age and spatial structure, but including heterogeneity in NPI adherence as well as disease structure, as illustrated in Supplementary Figure S5. Specifically, we assume a fraction f who regularly engage in NPI measures and are denoted by a ‡ (compartments $S^‡$, $I^‡$, $R_j^‡$ for $j = 1$ to n, which sum to f) and a fraction $1-f$ who do not (compartments S, I, R_j, which sum to $1-f$). See Appendix C for model details.

We first determine the benefit to an individual who adheres to NPI measures (e.g., masking). Near the endemic equilibrium, an individual engaging in the NPI measure has a lower risk of being infected at any given point in time of:

$$\underbrace{\frac{\hat{I}^‡/f}{\hat{I}/(1-f)}}_{\text{Relative risk of infection}} = \frac{1}{1+\hat{S}\,\frac{p}{(1-p)(1-f)}}, \tag{7}$$

where $\hat{S}$ is the fraction of the population at the endemic equilibrium who are susceptible and do not engage in the NPI measure (given by *Equation A11*, $\hat{S} \approx \kappa/\beta$ when f is small). This relative risk is mathematically equivalent to the relative rate at which individuals become infected for those who do

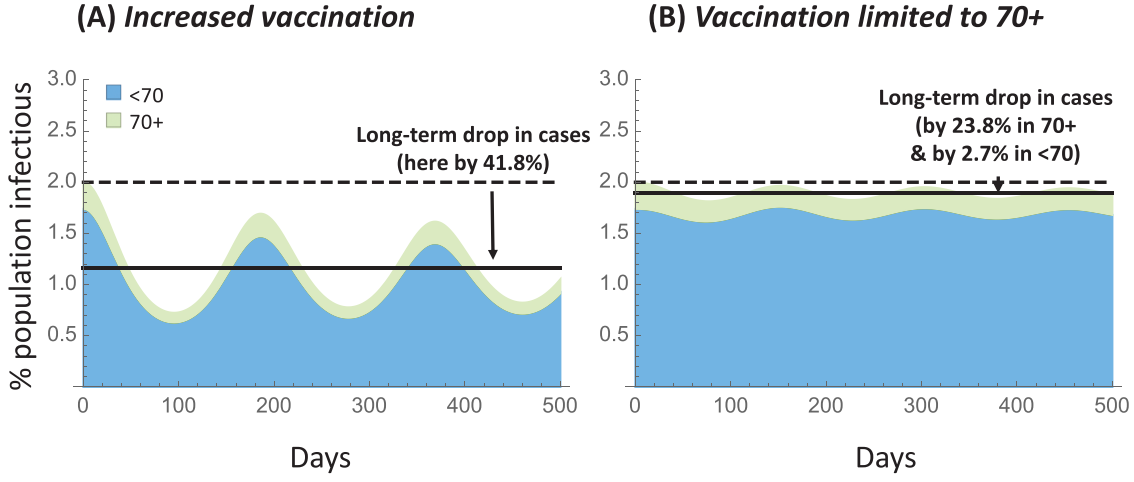

Figure 6. Impact of vaccination strategies. Plots illustrate the dynamics over time when vaccination is increased to 0.174% of the population per day (annual rate of 63.5%), either (A) within the entire population or (B) limited to a more vulnerable population (illustrated as 70 + in age). Parameters are identical to Figure 3, with $\nu = 0.00174$.

versus do not engage in the NPI measure, as well as the relative number of infections expected per year.

Figure 7 illustrates the relative risk of infection (*Equation 7*). As expected, the more effective the NPI measure is (the higher p) the lower the relative risk to individuals who engage in the NPIs (panel A). This individual-level benefit is only weakly dependent on the fraction of the population currently engaging in these measures (f = 10%, 50%, and 90% shown as solid, dashed, and dotted curves), with the relative risk rising slightly as f increases because non-practitioners gain a slight benefit from those who do practice the NPI measure.

The individual benefits depend strongly, however, on the endemic reproductive number of the disease ($\tilde{R}_0$). At the nominal value of $\tilde{R}_0 = 2$ and assuming low population-level uptake (f small), a person's relative risk of infection can be substantially reduced for NPIs that provide fairly modest protection (by 14% and by 33% for p = 25% and 50%, respectively). For $\tilde{R}_0 = 6$ (on the high end of the range considered plausible; Appendix A), however, these individual-level benefits diminish (to 5% and 14% for p = 25% and 50%, respectively), because individuals are exposed so often that modestly protective NPIs only moderately delay infection.

Not only is a susceptible individual who regularly engages in NPI measures less likely to become infected when in contact with an infectious individual (by a factor $1-p$), but they are also less likely to pass the infection on to one of their contacts (by another factor $1-p$), compared to a non-practicing individual who is currently susceptible. Accounting for the proportion of time that practicing and non-practicing individuals are susceptible (as in *Equation 7*), the risk per unit time of being infectious and infecting a contact is substantially lowered for those engaging in NPI measures relative to those who do not (Figure 7B). This validates the approach used by many who adhere to NPI measures, such as masking, in order to protect vulnerable contacts.

The curvature of the relative risks in Figure 7 highlights the utility of multiple complementary interventions: other policies that reduce transmission (lowering $\tilde{R}_0$) make masking more effective to individuals, because those individuals are less repeatedly exposed.

The individual-level benefits of NPI measures diminish with $\tilde{R}_0$ (x-axis of Figure 7) because individuals practicing NPI measures are more likely than non-practitioners to have remained uninfected and so are more often susceptible at the time of exposure, which increases their relative risk of infection as $\tilde{R}_0$ increases. Different results are obtained if individuals frequently switch their behavior (e.g., masking some days and not others). Modifying the model as described by *Equation A12*, all individuals are then equally likely to be susceptible on any given day, and the NPI measure always reduces the risk of infection by a factor $(1-p)$ for each practicing individual in an interaction. That is, the benefits remain at their maximal value of $(1-p)$ in panel A and $(1-p)^2$ in panel B, regardless of $\tilde{R}_0$ and f (Appendix C). Although not modeled, heterogeneity in contact rates is also likely to affect the value of NPI measures. If individuals who regularly adhere to NPI measures also strive to limit contact rates generally, their value of $\tilde{R}_0$ may be lower than non-practitioners. We conjecture, based on Figure 7, that NPI measures would benefit such individuals more than predicted from $\tilde{R}_0$ of the population as a whole.

We next evaluate the population-level advantages of NPI measures by calculating the fraction of infected individuals expected at the endemic equilibrium ($\hat{I}+\hat{I}^{\ddagger}$) when a fraction f of the population upholds these measures, relative to a population in which nobody does:

$$\underbrace{\frac{\hat{I}+\hat{I}^{\ddagger}}{\left(\hat{I}+\hat{I}^{\ddagger}\right)_{f=0}}}_{\text{Relative fraction of population infected}} = \left(1-\hat{S}-\hat{S}^{\ddagger}\right)\frac{\tilde{R}_0}{\tilde{R}_0-1}, \tag{8}$$

using the equilibrium values given by *Equation A11*. The right-hand side of *Equation 8* emphasizes that the population-wide benefits increase (fewer people will be infected) when there are more susceptible individuals available ($\hat{S}+\hat{S}^{\ddagger}$ larger).

While the population-level impact is small when few individuals engage in NPI measures (left panel of Figure 8), there are substantial benefits to having moderate to high adherence (central and right panels). These benefits are strongest when the endemic reproductive number is small, potentially moving the population away from the endemic case, where

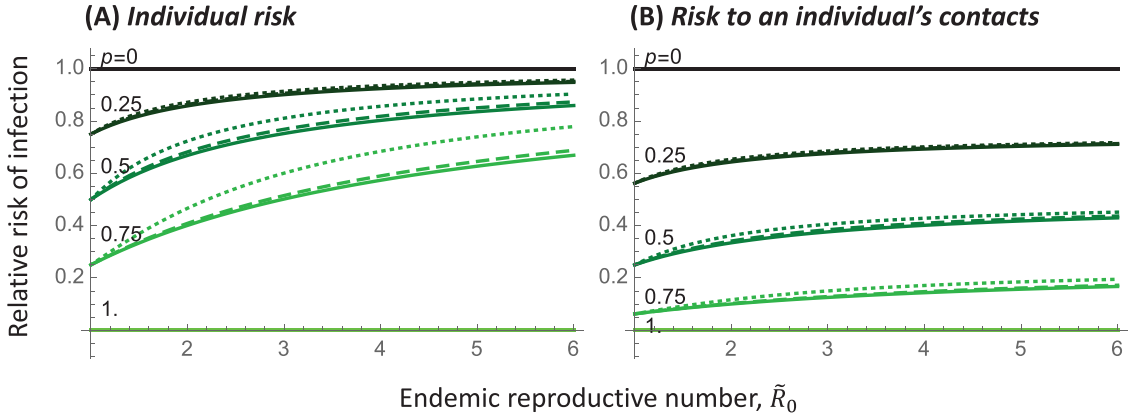

Figure 7. Risk of infection for individuals regularly engaging in an NPI measure such as masking, relative to unmasked individuals. Colored lines illustrate different levels of protection, p, provided by the NPI measure, in a population where the fraction of individuals engaging in the NPI measure is f = 10% (solid), 50% (dashed), and 90% (dotted). Panel A shows the risk of infection and panel B the risk of becoming infected and infecting a contact for an individual engaging in NPI measures, relative to those who do not. The x-axis gives the endemic reproductive number in this heterogeneous population, $\tilde{R}_0 = \left((1-f)+f(1-p)^2\right)\beta/\kappa$. Parameters: Relative risk depends on the parameters only through $\tilde{R}_0$, f, and p.

COVID-19 persists, to the disease-free equilibrium (when the curves cross the x-axis), again emphasizing the added benefits that come from combining interventions. While transmission-reducing measures are less valuable at a population level when exposure is very rapid ($\tilde{R}_0$ high), we estimate more modest exposure rates because of the substantial immunity in current populations (with $\tilde{R}_0$ ranging from {1.1,6.8}), making effective NPI measures still a valuable means of reducing population-wide incidence.

The reduction in cases caused by NPI measures is expected to result in a proportionate reduction in severe cases and deaths. Even a modest reduction (say 20%) can have nonlinear benefits when hospitals are at capacity, improving care for all (Wichmann & Wichmann, 2023). Achieving such benefits at a population level, however, requires that there be clear messaging and incentives to obtain the moderate to high levels of uptake required to impact population-wide infection rates.

Discussion

This article aims to expand our understanding of the impact of variants, as well as behavioral and public health measures, on endemic diseases like COVID-19. Widespread measures, both by individuals and public health agencies, repeatedly "flattened the curve" of COVID-19 during the first 2 years of the pandemic, reducing viral transmission to save lives and avoid the collapse of healthcare systems (Ogden et al., 2022; Talic et al., 2021). Since mid-2022, however, COVID-19 has persisted at high levels throughout the world, becoming endemic with no sign of abating even during the summer months. The mantra to "flatten the curve" is no longer relevant, as endemic levels are already fairly flat, and we lack a compelling guide to govern our collective behavior in its place.

For COVID-19, endemic does not mean constant, with wavelets expected as a result of new variants, changing behavior, vaccination campaigns, as well as damped oscillations from previous waves ("echo" waves). Nor does endemic mean rare, as ongoing high levels of COVID-19 health impacts remain. Nor does endemic mean out of our control, as protective measures continue to have important benefits, boosting immunity through vaccination and reducing transmission through effective NPI measures. The goal of this article is twofold: to explore the impact of evolutionary changes in the virus on disease incidence and to discuss how protective measures can counteract these rises, reducing disease risks.

Variants of endemic diseases that increase transmissibility and/or immune evasion are selectively favored, with rises in frequency that can be measured empirically, yielding estimates of the strength of selection (s). While the strength of selection accurately predicts the speed with which one variant replaces another, it does not predict the long-term impact on endemic levels of disease. For a given selection coefficient, we have shown that the long-term impact on disease is negligible for a variant that is more immune evasive, but only transiently so, eliciting variant-specific antibodies that protect from reinfection (Figure 5A). By contrast, an immune evasive variant that fails to elicit variant-specific antibodies has a persistent advantage, leading to a nearly proportional increase in cases in the long term (Figure 5B). In Appendix B, we also consider a variant that causes immunity to become leakier, increasing the risk of infection for all recovered classes, which are particularly problematic (Supplementary Figure S3), causing a high long-term rise in cases because all individuals remain prone to infection if leakiness is persistent. A variant that is more transmissible generally has an intermediate impact on disease incidence (Figure 5C). Thus, depending on the exact properties of a new variant, we may see smaller or larger rises in cases over the long term, even for variants initially spreading at the same rate. Of course, a series of variants can lead to continual short-term rises in cases, but the analyses conducted here allow the long-term effects of each variant to be assessed separately.

As multiple variants arise and spread, their impact on the incidence of disease is predicted to accumulate (Supplementary Figure S4). For persistent immune evasive variants, however, subsequent infections with similar variants are expected to improve the breadth of neutralizing antibodies, gradually overriding immune imprinting and improving variant specific responses (Yisimayi et al., 2023). The co-evolution of variants and variant-specific immunity could be explicitly modeled in future work by lowering the degree of immune evasiveness (m) with subsequent infections. It would also be valuable to couple the expected changes in incidence predicted here to the rate of appearance of variants, as mutations continue to accumulate in the large global population of infected individuals.

Lab assays of SARS-CoV-2 have dramatically sped up phenotypic assessment of new variants (Cao et al., 2023). Within days of new variants emerging, information has been shared by groups around the world, evaluating immune evasiveness (e.g., the titer of neutralizing antibodies in convalescent

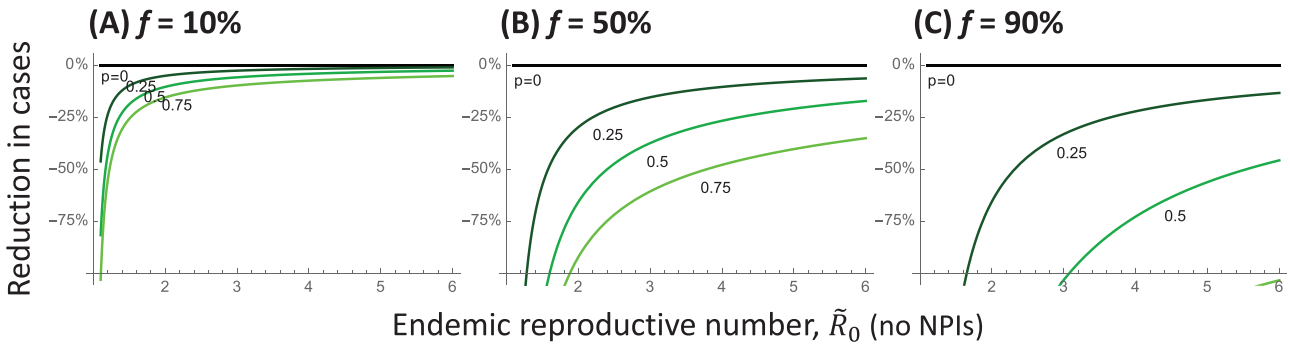

Figure 8. Reduction in cases at the endemic equilibrium when a fraction f of the population engages in an NPI measure, relative to when none do. Colored curves illustrate different levels of protection, p, provided by the NPI measure. Panels show the fraction of the population practicing the NPI measure: (A) f = 10%, (B) 50%, and (C) 90%. The x-axis gives the endemic reproductive number in a population that is not engaging in the NPI measure, given by $\tilde{R}_0 = \beta/\kappa$. Parameters: Reduction in cases depends on the parameters only through $\tilde{R}_0$, f, and p.

plasma required to prevent infection of cell lines) and efficiency of binding to ACE-2 receptors (e.g., via twitter.com, @yunlong_cao). These assays often find that infection with one variant (e.g., BA.1) builds higher neutralizing capacity against that variant than other variants (e.g., BA.2), indicating some loss of immune evasiveness following infection with a variant (Cao et al., 2023). The impact on long-term immune evasion and reinfection rates for different variants remains an open question, and one whose answer determines the impact on endemic incidence of disease (Figure 5A and B).

We can counter variant-induced rises in cases, however, by encouraging higher uptake rates of vaccines and other non-pharmaceutical interventions. These measures always help individuals reduce their own risk of infection and the risk of infecting those around them (Figure 7). Widespread, but not universal, uptake is needed to substantially reduce levels of disease (Figure 8), except if the disease is near eradication ($\tilde{R}_0$ near 1). The benefits could be enhanced by encouraging NPI measures around those who are most at risk of adverse outcomes and in places and times where risks of infection and/or the health care burden are high. Particularly valuable are investments in measures that protect all, regardless of uptake (such as improved air filtration and ventilation, adequate testing and job security to stay home when sick).

The models explored herein lack many important epidemiological details, including spatial and age structure in contact rates and seasonal variation in transmission risk. As such, the results are meant to guide expectations rather than provide precise predictions. Details were sacrificed in an effort to help us better understand how the endemic level of disease is likely to change in the future, in response to our efforts as well as further evolution of the virus.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

All proofs and code needed to generate the figures are available in *Mathematica* and PDF versions on Zenodo (https://zenodo.org/records/10783773). *Mathematica* files can be read using Wolfram's *Mathematica* player (https://www.wolfram.com/player/) or re-analyzed with a low-cost version of *Mathematica* using a Raspberry Pi installation (https://projects.raspberrypi.org/en/projects/getting-started-with-mathematica).

Funding

Funding was provided by the Natural Sciences and Engineering Research Council of Canada to S.P.O. (RGPIN-2022-03726), to A.M. (RGPIN-2022-03113), and to C.C. (CANMOD; RGPIN-2019-06624), by the Canada Research Chairs Program (S.P.O. and A.M.) and the Canada 150 Research Chair Program (C.C.), and by the Canadian Institutes for Health Research (CIHR) operating grant to CoVaRR-Net.

Conflict of interest: The authors declare no competing interests.

Acknowledgments

We thank all the authors, developers, and contributors to the VirusSeq Data Portal (https://virusseq-dataportal.ca/) for making their SARS-CoV-2 sequences publicly available. We wish to acknowledge the Canadian Public Health Laboratory Network (CPHLN), Genome Canada and the CanCOGeN VirusSeq Consortium for their contributions to the Portal, see supplementary file for detailed information. The variant calls from the Portal were used for Canadian SARS-CoV-2 samples were used to generate Figure 1. The authors would like to thank the scientific community of the Coronavirus Variants Rapid Response Network (CoVaRR-Net), particularly Fiona Brinkman, Carmen Lia Murall, Jesse Shapiro, Erin Gill, and Justin Jia for helpful discussions about variants. We are grateful for the comments received on the models and the paper from Cyrus Bhiladvala, Joanna Masel, Tim Connallon, and an anonymous reviewer.

References

Aghaali, M., Kolifarhood, G., Nikbakht, R., Saadati, H. M., & Nazari, S. S. H. (2020). Estimation of the serial interval and basic reproduction number of COVID-19 in Qom, Iran, and three other countries: A data-driven analysis in the early phase of the outbreak. *Transboundary and Emerging Diseases*, *67*(6), 2860–2868.

Anderson, S. C., Edwards, A. M., Yerlanov, M., Mulberry, N., Stockdale, J. E., Iyaniwura, S. A., Falcao, R. C., Otterstatter, M. C., Irvine, M. A., Janjua, N. Z., Coombs, D., & Colijn, C. (2020). Quantifying the impact of COVID-19 control measures using a Bayesian model of physical distancing. *PLoS Computational Biology*, *16*(12), e1008274. https://doi.org/10.1371/journal.pcbi.1008274

Andrews, N., Stowe, J., Kirsebom, F., Toffa, S., Rickeard, T., Gallagher, E., Gower, C., Kall, M., Groves, N., O'Connell, A. -M., Simons, D., Blomquist, P. B., Zaidi, A., Nash, S., Iwani Binti Abdul Aziz, N., Thelwall, S., Dabrera, G., Myers, R., Amirthalingam, G., … Lopez Bernal, J. (2022). Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *The New England Journal of Medicine*, *386*(16), 1532–1546. https://doi.org/10.1056/NEJMoa2119451

Arabi, M., Al-Najjar, Y., Sharma, O., Kamal, I., Javed, A., Gohil, H. S., Paul, P., Al-Khalifa, A. M., Laws, S. 'ad, & Zakaria, D. (2023). Role of previous infection with SARS-CoV-2 in protecting against omicron reinfections and severe complications of COVID-19 compared to pre-omicron variants: A systematic review. *BMC Infectious Diseases*, *23*(1), 432. https://doi.org/10.1186/s12879-023-08328-3

Canadian Blood Services. (2023). *COVID-19 Seroprevalence report July 28, 2023*. https://www.covid19immunitytaskforce.ca/wp-content/uploads/2023/08/covid-19-full-report-june-2023.pdf

Cao, Y., Jian, F., Wang, J., Yu, Y., Song, W., Yisimayi, A., Wang, J., An, R., Chen, X., Zhang, N., Wang, Y., Wang, P., Zhao, L., Sun, H., Yu, L., Yang, S., Niu, X., Xiao, T., Gu, Q., … Xie, X. S. (2023). Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *Nature*, *614*(7948), 521–529. https://doi.org/10.1038/s41586-022-05644-7

CDC. (2023). *Stay up to date with COVID-19 vaccines*. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html

Chemaitelly, H., Ayoub, H. H., AlMukdad, S., Coyle, P., Tang, P., Yassine, H. M., Al-Khatib, H. A., Smatti, M. K., Hasan, M. R., Al-Kanaani, Z., Al-Kuwari, E., Jeremijenko, A., Kaleeckal, A. H., Latif, A. N., Shaik, R. M., Abdul-Rahim, H. F., Nasrallah, G. K., Al-Kuwari, M. G., Butt, A. A., … Abu-Raddad, L. J. (2022). Duration of MRNA vaccine protection against SARS-CoV-2 Omicron BA.1 and BA.2 subvariants in Qatar. *Nature Communications*, *13*(1), 3082. https://doi.org/10.1038/s41467-022-30895-3

CoVaRR-Net's CAMEO. (2023). *Duotang, A genomic epidemiology analyses and mathematical modelling notebook*. https://covarr-net.github.io/duotang/duotang.html

COVID-19 Resources Canada. (2023). *Detailed forecasts: CAN, followed by each province on subsequent pages*. https://covid19resources.ca/covid-hazard-index/

Day, T., Gandon, S., Lion, S., & Otto, S. P. (2020). On the evolutionary epidemiology of SARS-CoV-2. *Current Biology*, *30*(15), R849–R857.

Eikenberry, S. E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y., Kostelich, E., & Gumel, A. B. (2020). To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling*, *5*, 293–308. https://doi.org/10.1016/j.idm.2020.04.001

Erikstrup, C., Laksafoss, A. D., Gladov, J., Kaspersen, K. A., Mikkelsen, S., Hindhede, L., Boldsen, J. K., Jørgensen, S. W., Ethelberg, S., Holm, D. K., Bruun, M. T., Nissen, J., Schwinn, M., Brodersen, T., Mikkelsen, C., Sækmose, S. G., Sørensen, E., Harritshøj, L. H., Aagaard, B., … Pedersen, O. B. V. (2022). Seroprevalence and infection fatality rate of the SARS-CoV-2 Omicron variant in Denmark: A nationwide serosurveillance study. *The Lancet Regional Health. Europe*, *21*, 100479. https://doi.org/10.1016/j.lanepe.2022.100479

Evans, J. P., Zeng, C., Carlin, C., Lozanski, G., Saif, L. J., Oltz, E. M., Gumina, R. J., & Liu, S. -L. (2022). Neutralizing antibody responses elicited by SARS-CoV-2 MRNA vaccination wane over time and are boosted by breakthrough infection. *Science Translational Medicine*, *14*(637), eabn8057. https://doi.org/10.1126/scitranslmed.abn8057

Health Infobase Canada. (2023). *COVID-19 vaccination: Doses administered*. https://health-infobase.canada.ca/covid-19/vaccine-administration/

Hethcote, H. W., Stech, H. W., & Van Den Driessche, P. (1981). Nonlinear oscillations in epidemic models. *SIAM Journal on Applied Mathematics*, *40*(1), 1–9. https://doi.org/10.1137/0140001

Jacobsen, H., Sitaras, I., Katzmarzyk, M., Jimenez, V. C., Naughton, R., Higdon, M. M., & Knoll, M. D. (2023). Systematic review and meta-analysis of the factors affecting waning of post-vaccination neutralizing antibody responses against SARS-CoV-2. *NPJ Vaccines*, *8*, 159. https://doi.org/10.1038/s41541-023-00756-1

Ke, R., Romero-Severson, E., Sanche, S., & Hengartner, N. (2021). Estimating the reproductive number R0 of SARS-CoV-2 in the United States and eight European countries and implications for vaccination. *Journal of Theoretical Biology*, *517*, 110621. https://doi.org/10.1016/j.jtbi.2021.110621

Keeling, M. J., & Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.

Khoury, D. S., Cromer, D., Reynaldi, A., Schlub, T. E., Wheatley, A. K., Juno, J. A., Subbarao, K., Kent, S. J., Triccas, J. A., & Davenport, M. P. (2021). Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature Medicine*, *27*(7), 1205–1211. https://doi.org/10.1038/s41591-021-01377-8

Lau, C. S., Oh, M. L. H., Phua, S. K., Liang, Y. -L., & Aw, T. C. (2022). 210-day kinetics of total, IgG, and neutralizing spike antibodies across a course of 3 doses of BNT162b2 MRNA vaccine. *Vaccines*, *10*(10), 1703. https://doi.org/10.3390/vaccines10101703

Lau, E. H., Hui, D. S., Tsang, O. T., Chan, W. -H., Kwan, M. Y., Chiu, S. S., Cheng, S. M., Ko, R. L., Li, J. K., Chaothai, S., Tsang, C. H., Poon, L. L., & Peiris, M. (2021). Long-term persistence of SARS-CoV-2 neutralizing antibody responses after infection and estimates of the duration of protection. *EClinicalMedicine*, *41*, 101174. https://doi.org/10.1016/j.eclinm.2021.101174

Leech, G., Rogers-Smith, C., Monrad, J. T., Sandbrink, J. B., Snodin, B., Zinkov, R., Rader, B., Brownstein, J. S., Gal, Y., Bhatt, S., Sharma, M., Mindermann, S., Brauner, J. M., & Aitchison, L. (2022). Mask wearing in community settings reduces SARS-CoV-2 transmission. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(23), e2119266119. https://doi.org/10.1073/pnas.2119266119

Li, H., Yuan, K., Sun, Y. -K., Zheng, Y. -B., Xu, Y. -Y., Su, S. -Z., Zhang, Y. -X., Zhong, Y., Wang, Y. -J., Tian, S. -S., Gong, Y. -M., Fan, T. -T., Lin, X., Gobat, N., Wong, S. Y. S., Chan, E. Y. Y., Yan, W., Sun, S. -W., Ran, M. -S., … Lu, L. (2022). Efficacy and practice of facemask use in general population: A systematic review and meta-analysis. *Translational Psychiatry*, *12*(1), 49. https://doi.org/10.1038/s41398-022-01814-3

Lin, C., Bier, B., Tu, R., Paat, J. J., & Tu, P. (2023). Vaccinated yet booster-hesitant: Perspectives from boosted, non-boosted, and unvaccinated individuals. *Vaccines*, *11*(3), 550. https://doi.org/10.3390/vaccines11030550

Lind, M. L., Dorion, M., Houde, A. J., Lansing, M., Lapidus, S., Thomas, R., Yildirim, I., Omer, S. B., Schulz, W. L., Andrews, J. R., Hitchings, M. D. T., Kennedy, B. S., Richeson, R. P., Cummings, D. A. T., & Ko, A. I. (2023). Evidence of leaky protection following COVID-19 vaccination and SARS-CoV-2 infection in an incarcerated population. *Nature Communications*, *14*(1), 5055. https://doi.org/10.1038/s41467-023-40750-8

Mahase, E. (2023). Covid-19: Annual flu-like booster approach may not be appropriate, says expert on infectious disease. *BMJ*, *380*, 196. https://doi.org/10.1136/bmj.p196

Menegale, F., Manica, M., Zardini, A., Guzzetta, G., Marziano, V., d'Andrea, V., Trentini, F., Ajelli, M., Poletti, P., & Merler, S. (2023). Evaluation of waning of SARS-CoV-2 vaccine-induced immunity: A systematic review and meta-analysis. *JAMA Network Open*, *6*(5), e2310650. https://doi.org/10.1001/jamanetworkopen.2023.10650

Mulberry, N., Tupper, P., Kirwin, E., McCabe, C., & Colijn, C. (2021). Vaccine rollout strategies: The case for vaccinating essential workers early. *PLOS Global Public Health*, *1*(10), e0000020. https://doi.org/10.1371/journal.pgph.0000020

NACI. (2023). *Guidance on the use of COVID-19 vaccines in the fall of 2023*. Public Health Agency of Canada. https://www.canada.ca/content/dam/phac-aspc/documents/services/publications/vaccines-immunization/national-advisory-committee-immunization-guidance-use-covid-19-vaccines-fall-2023/statement.pdf

NHS. (2023). *About COVID-19 vaccination*. https://www.nhs.uk/conditions/covid-19/covid-19-vaccination/about-covid-19-vaccination/

Nyberg, T., Ferguson, N. M., Nash, S. G., Webster, H. H., Flaxman, S., Andrews, N., Hinsley, W., Bernal, J. L., Kall, M., Bhatt, S., Blomquist, P., Zaidi, A., Volz, E., Aziz, N. A., Harman, K., Funk, S., Abbott, S., Hope, R., Charlett, A., … Thelwall, S. (2022). Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: A cohort study. *The Lancet*, *399*(10332), 1303–1312. https://doi.org/10.1016/s0140-6736(22)00462-7

Office for National Statistics. (2023). *Results—Weekly swab positivity updates from ONS*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/24march2023

Ogden, N. H., Turgeon, P., Fazil, A., Clark, J., Gabriele-Rivet, V., Tam, T., & Ng, V. (2022). Counterfactuals of effects of vaccination and public health measures on COVID-19 cases in Canada: What could have happened? *Canada Communicable Disease Report*, *48*(7–8), 292–302. https://doi.org/10.14745/ccdr.v48i78a01

Otto, S. P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., Mechai, S., Van Domselaar, G., Wu, J., Earn, D. J. D., & Ogden, N. H. (2021). The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Current Biology*, *31*(14), R918–R929.

Our World in Data. (2023). *Daily COVID-19 vaccine doses administered*. https://ourworldindata.org/grapher/daily-covid-19-vaccination-doses

Oved, K., Olmer, L., Shemer-Avni, Y., Wolf, T., Supino-Rosin, L., Prajgrod, G., Shenhar, Y., Payorsky, I., Cohen, Y., Kohn, Y., Indenbaum, V., Lazar, R., Geylis, V., Oikawa, M. T., Shinar, E., Stoyanov, E., Keinan-Boker, L., Bassal, R., Reicher, S., … Lustig, Y. (2020). Multi-center nationwide comparison of seven serology assays reveals a SARS-CoV-2 non-responding seronegative subpopulation. *EClinicalMedicine*, *29*, 100651. https://doi.org/10.1016/j.eclinm.2020.100651

Puhach, O., Adea, K., Hulo, N., Sattonnet, P., Genecand, C., Iten, A., Jacquérioz, F., Kaiser, L., Vetter, P., Eckerle, I., & Meyer, B. (2022). Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron SARS-CoV-2. *Nature Medicine*, *28*(7), 1491–1500. https://doi.org/10.1038/s41591-022-01816-0

The Royal Society. (2023). *COVID-19: Examining the effectiveness of non-pharmaceutical interventions*. https://royalsociety.org/-/media/policy/projects/impact-non-pharmaceutical-interventions-on-covid-19-transmission/

the-royal-society-covid-19-examining-the-effectiveness-of-non-pharmaceutical-interventions-report.pdf
Scherer, A., & McLean, A. (2002). Mathematical models of vaccination. *British Medical Bulletin*, *62*, 187–199. https://doi.org/10.1093/bmb/62.1.187
Statistics Canada. (2023). *Population estimates on July 1st, by age and sex*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501
Talic, S., Shah, S., Wild, H., Gasevic, D., Maharaj, A., Ademi, Z., Li, X., Xu, W., Mesa-Eguiagaray, I., Rostron, J., Theodoratou, E., Zhang, X., Motee, A., Liew, D., & Ilic, D. (2021). Effectiveness of public health measures in reducing the incidence of Covid-19, SARS-CoV-2 transmission, and Covid-19 mortality: Systematic review and meta-analysis. *BMJ*, *375*, e068302. https://doi.org/10.1136/bmj-2021-068302
Tan, S. T., Kwan, A. T., Rodríguez-Barraquer, I., Singer, B. J., Park, H. J., Lewnard, J. A., Sears, D., & Lo, N. C. (2023). Infectiousness of SARS-CoV-2 breakthrough infections and reinfections during the Omicron wave. *Nature Medicine*, *29*(2), 358–365. https://doi.org/10.1038/s41591-022-02138-x
UKHSA. (2023). *COVID-19 Omicron variant infectious period and transmission from people with asymptomatic compared with symptomatic infection: A rapid review*. GOV-14430. UK Health Security Agency. https://www.gov.uk/government/publications/covid-19-omicron-variant-infectious-period-and-asymptomatic-and-symptomatic-transmission
van Dorp, C. H. van, Goldberg, E. E., Hengartner, N., Ke, R., & Romero-Severson, E. O. (2021). Estimating the strength of selection for new SARS-CoV-2 variants. *Nature Communications*, *12*(1), 7239. https://doi.org/10.1038/s41467-021-27369-3
VirusSeq. (2023). *Canadian Virus Seq data portal*. https://virusseq-dataportal.ca/
Wei, J., Stoesser, N., Matthews, P. C., Ayoubkhani, D., Studley, R., Bell, I., Bell, J. I., Newton, J. N., Farrar, J., Diamond, I., Rourke, E., Howarth, A., Marsden, B. D., Hoosdally, S., Jones, E. Y., Stuart, D. I., Crook, D. W., Peto, T. E. A., Pouwels, K. B., … Walker, A. S.; COVID-19 Infection Survey team (2021). Antibody responses to SARS-CoV-2 vaccines in 45,965 adults from the general population of the United Kingdom. *Nature Microbiology*, *6*(9), 1140–1149. https://doi.org/10.1038/s41564-021-00947-3
Wichmann, B., & Wichmann, R. M. (2023). Big data evidence of the impact of COVID-19 hospitalizations on mortality rates of non-COVID-19 critically ill patients. *Scientific Reports*, *13*(1), 13613. https://doi.org/10.1038/s41598-023-40727-z
Xin, H., Wang, Z., Feng, S., Sun, Z., Yu, L., Cowling, B. J., Kong, Q., & Wu, P. (2023). Transmission dynamics of SARS-CoV-2 Omicron variant infections in Hangzhou, Zhejiang, China, January-February 2022. *International Journal of Infectious Diseases*, *126*, 132–135. https://doi.org/10.1016/j.ijid.2022.10.033
Yisimayi, A., Song, W., Wang, J., Jian, F., Yu, Y., Chen, X., Xu, Y., Yang, S., Niu, X., Xiao, T., Wang, J., Zhao, L., Sun, H., An, R., Zhang, N., Wang, Y., Wang, P., Yu, L., Lv, Z., … Cao, Y. (2023). Repeated Omicron exposures override ancestral SARS-CoV-2 immune imprinting. *Nature*, *625*(7993), 148–156. https://doi.org/10.1038/s41586-023-06753-7
Zhou, Z., Barrett, J., & He, X. (2023). Immune imprinting and implications for COVID-19. *Vaccines*, *11*(4), 875. https://doi.org/10.3390/vaccines11040875

Appendix A: Modeling the spread of variants

Dynamics

We include infections by a resident variant (I) and a new variant (I^*) in the SIR$_n$ epidemiological model illustrated in Figure 1. By allowing for multiple recovered classes, we can model new variants that are more immune evasive by allowing them to infect earlier in the waning period (infecting individuals in the last m recovered compartments R_j), when antibody levels are high enough to prevent infection by the resident virus but not the new variant. The dynamics are then described by the following set of differential equations:

$$\frac{dS}{dt} = \delta_n R_n - \beta S I - \beta^* S I^*$$

$$\frac{dI}{dt} = \beta S I - \kappa I \quad \frac{dI^*}{dt} = \beta^* \left(S + \sum_{j=1+n-m}^{n} R_j \right) I^* - \kappa I^* \tag{A1}$$

$$\frac{dR_1}{dt} = \kappa I + \kappa I^* - \delta_1 R_1$$

$$\frac{dR_j}{dt} = \delta_{j-1} R_{j-1} - \delta_j R_j \quad \text{for} \quad 2 \le j \le n-m$$

$$\frac{dR_j}{dt} = \delta_{j-1} R_{j-1} - \delta_j R_j - \beta^* R_j I^* \quad \text{for } n-m < j \le n$$

Setting all waning rates between recovered classes equal to $\delta_i = \delta/n$ ensures that the average time from first recovering to returning to the susceptible class has a mean of $1/\delta$ days. The distribution of waning times is then given by a gamma distribution with a coefficient of variation (CV) of $1/\sqrt{n}$, becoming more bell shaped with higher n (Hethcote et al., 1981).

Spread of a new variant

The spread of a new variant into a population at the stable endemic equilibrium (*Equation 2*) is given by the leading eigenvalue, λ_L of the external stability matrix describing the dynamics of the variant (see details in the Supplementary *Mathematica* package), which equals:

$$\lambda_L = \hat{S}\beta^* + m\hat{R}_j\beta^* - \kappa \tag{A2}$$

If the new variant did not change the transmission rate ($\beta^* = \beta$) and was unable to infect any additional sector of the population ($m = 0$), it would be neutral ($\lambda_L = 0$, plugging in *Equation 2*).

The selection coefficient favoring a new variant is defined by the rise in frequency of the new variant relative to the old variant ($\frac{dx}{dt} \equiv s\,x$, where x = freq(new variant)/freq(old variant)), which predicts an exponential rise in the relative frequency of the new variant over time ($x_t = e^{s\,t} x_0$). The strength of selection can thus be estimated empirically by the slope on a logit plot (plotting log of x_t over time). Near the endemic equilibrium, it can be shown that selection, defined in this way, equals λ_L (see Supplementary *Mathematica* package). Plugging in *Equation 2* for $\hat{S}$ into *Equation A2* then gives the selection coefficient reported in *Equation 4*.

Parameter values

We consider the following parameter values for the current endemic phase during which Omicron predominates, giving the nominal value considered typical and the plausible range in square brackets:

- κ of 0.2 (mean of 5 days) [range of 3–10 days]. Source: Estimates of the infectious period for Omicron vary depending on the study design, but several studies are consistent with infectiousness for a couple of days prior to symptom onset and 5 days thereafter (UKHSA, 2023). We take into account some self-isolation upon infection and use a 5-day average infectious period as a default.

- δ of 0.008 (mean of 125 days) [range of 100–180 days]. Source: Waning rate depends on the exact sequence of vaccinations and infections. The half-life of protection against symptomatic infection with Omicron among studies summarized by Menegale et al. (Menegale et al., 2023) was 87 days without a booster and 111 days with a booster, yielding δ values ranging from 0.0071 to 0.0094 per day. Waning rates were similar for older and younger individuals (Figure 14 of Menegale et al., 2023).
- $\hat{I}$ of 2% [range of 0.5%–4%]. Sources: The last report of the Coronavirus (COVID-19) Infection Survey UK (Office for National Statistics, 2023), which assayed nose and throat swabs from households, found 2.66% of England were infected (March 13, 2023). In Canada, models suggest that 1 in 28 were infected the week of April 16, 2023, while 1 in 80 were infected the week of 9 September 2023 (COVID-19 Resources Canada, 2023).
- Age structure for Canada: 13% of the population is 70 + in age (Statistics Canada, 2023).

Combining these estimates with *Equations 2 and 3* allows estimation of the transmission rate and reproductive number. The nominal parameters given above yield estimates of $\beta = 0.42$ and $\tilde{R}_0 = 2.1$, ranging from $\beta = \{0.11–2.27\}$ and $\tilde{R}_0 = \{1.1,6.8\}$ (Supplementary Table S1), although some combinations are not possible (e.g., a mean waning time of 180 days is inconsistent with an incidence of $\hat{I} = 4\%$ if mean clearance times are too short, $\kappa > 0.13$), and the disease is then expected to decline.

The expected number of disease bouts per year is 1.46 for the nominal parameter values, ranging from 0 (when the disease disappears) to 2.92 when incidence is high ($\hat{I} = 4\%$), waning is fast ($\delta = 1/100$), and recovery is fast but not so fast that the disease disappears ($\kappa = 1/5$).

We can also calculate the expected number of infections per year for an individual who is vaccinated at regular intervals (every T days). For simplicity, we make the approximation that vaccinations are frequent enough and waning slow enough that we need only consider the chance of one infection between vaccinations. If waning times were exponentially distributed, then the probability of becoming infected in the period between vaccinations would be:

$$P = \int_0^T \underbrace{\delta\, e^{-\delta t}}_{\text{Waning at time t}} * \underbrace{\left(1 - e^{-\beta \hat{I}(T-t)}\right)}_{\text{Probability of infection after waning}} dt \tag{A3}$$

$$= 1 - \frac{e^{-\delta T}\beta\hat{I} - e^{-\beta\hat{I}T}\delta}{\beta\hat{I} - \delta}.$$

The approximate annual number of infections is then $365\ P/T$, which is 0.88 for those on a 6-month vaccination interval ($T = 365/2$) and the nominal parameter values ($\kappa = 0.2, \delta = 0.008, \hat{I} = 2\%, \beta = 0.42$).

Appendix B: Model sensitivity and oscillatory behavior

Different choices about the number of recovered classes, movement among them, and whether immunity is leaky, as well as the inclusion of a latent period and incomplete seroconversion, were explored to determine sensitivity of the results to model assumptions (Supplementary *Mathematica* package). To simplify the presentation, we focus on the case where daily vaccination rates are low and are ignored (except where noted).

Alternate models of recovery

In the main text, we used multiple recovered classes in the SIR_n model to capture observed declines in neutralizing antibodies, measured on a log scale, over time since vaccination and/or infection. While this reflects the dynamics of neutralizing antibody levels, a side consequence is that the distribution for the total waning time becomes increasingly bell shaped as n rises (CV of $1/\sqrt{n}$). This synchronizes the recovery of individuals infected at the same time. If n is large enough, this synchronization can destabilize the endemic equilibrium, leading to persistent cycles (Hethcote et al., 1981). While the rise in frequency of a variant, as described by its selective advantage (*Equation 4*) and the long-term impact of the variant on the endemic equilibrium (*Equation 5*) are insensitive to the number of recovered compartments (n), the extent of oscillations following the initial spread of the variant are much stronger as n increases (Supplementary Figure S2, panels A–C). Empirically, the distribution of waning times is close to exponential (CV = 1; Menegale et al., 2023), suggesting that intrinsic oscillations are likely to be damped (like Supplementary Figure S2A, where CV = 1).

Similar behaviors to Supplementary Figure S2A are seen in a model with only two recovered classes ($n = 2$), corresponding to high (R_1) and low (R_2) antibody levels, where an immune evasive variant (but not the resident virus) can infect the second class. By setting the waning rates to $\delta_1 = \delta/x$ (from R_1 to R_2) and $\delta_2 = \delta/(1-x)$ (from R_2 to S), the equilibrium fraction of recovered individuals in the second class (x) can be adjusted to allow for more immune evasive variants, while keeping the average time from first recovering to susceptibility at $1/\delta$ days for the resident virus. This model has a nearly exponential waning time, with rapidly dampening oscillations, for more transmissible variants (Supplementary Figure S2D), more immune evasive variants (Supplementary Figure S2E), or both (Supplementary Figure S2F). The selection coefficient and equilibrium remain unchanged, all else being equal (given by *Equations 4 and 5*, respectively).

Leaky immunity

In the main text, we considered immunity to be polarized: individuals are either susceptible to infection (S compartment) or not (R_j compartments, with j depending on the variant). There is evidence, however, that SARS-CoV-2 immunity is leaky, such that high viral exposure can lead to infection for those who would otherwise be immune (Lind et al., 2023). Furthermore, variants may differ in the extent of leaky immunity (e.g., Lind et al., 2023 found higher hazard ratios following close exposure for Delta than for Omicron).

We thus explored variants that increased leakiness of immunity, ξ, in the SIR ($n = 1$) and SIR_n ($n = 5$) models (exploring the latter numerically only in the Supplementary *Mathematica* package). Incorporating leaky immunity in the SIR model changes the dynamics to:

$$\frac{dS}{dt} = \delta\ R - \beta SI - \beta^* SI^*$$

$$\begin{aligned}\frac{dI}{dt} &= \beta SI + \xi\beta RI - \kappa I\\ \frac{dI^*}{dt} &= \beta^* S\, I^* + \xi^*\beta^* R\, I^* - \kappa\, I^*\\ \frac{dR}{dt} &= \kappa\, I + \kappa\, I^* - \xi\beta RI - \xi^*\beta^* R\, I^* - \delta\, R\end{aligned} \tag{A4}$$

The equilibrium is then:

$$\hat{S} = \frac{1}{2}\left(-b + \sqrt{b^2 - 4c}\right)$$

$$\hat{I} = \frac{\delta\,\left(\frac{\kappa}{\beta} - \hat{S}\right)}{\hat{S}\xi\beta} \tag{A5}$$

where $b = -\frac{\kappa\, -\delta-\xi\beta}{(1-\xi)\beta}$ and $c = -\frac{\kappa}{\beta}\,\frac{\delta}{(1-\xi)\beta}$. Selection on a variant then becomes:

$$s = \underbrace{\frac{\Delta\beta}{\beta}\,\kappa}_{\text{Transmission advantage}} + \underbrace{\Delta\xi\;\hat{R}\;\beta^*}_{\text{Evasion}\quad\text{advantage}} \tag{A6}$$

Supplementary Figure S3 illustrates cases where immunity was robust against the resident virus ($\xi = 0$) but leaky for the variant ($\xi^* > 0$), combined with some to no transmission advantage (panels A–C). Again we see that the same selective advantage (s) is consistent with substantially different long-term consequences for endemic disease levels. Variants that exhibit leakier immunity greatly increase the endemic equilibrium, more than seen in Figures 3 and 4 for a given selection coefficient, because all individuals are more prone to infection in the long term, not just those with low antibody levels.

Latent period

Viral infections are characterized by a latent period between infection and detectible viral load, which is thought to indicate the onset of the infectious period (UKHSA, 2023). We can include this period by adding to the SIR_n model a latent class (E), into which all new infections enter and then exit at rate ϵ. Including this period, the equilibrium fraction of infected individuals changes to:

$$\hat{I} = \left(1 - \frac{\kappa}{\beta}\right)\frac{\delta\;\epsilon}{\epsilon\;\delta + \epsilon\;\kappa + \delta\;\kappa} \tag{A7}$$

Given that the rate of leaving the latent class is much faster than waning ($\epsilon \gg \delta$), the last $\delta\;\kappa$ term in the denominator is negligible, and ϵ cancels out of *Equation A7*. Thus, the equilibrium number of infections $\left(\hat{I}\right)$ is nearly unaffected by including a latent class.

Recalculating the leading eigenvalue at this endemic equilibrium, the selection coefficient favoring the new variant ($s = \lambda_L$) changes slightly when a latent period is added, from s given by *Equation 4* to:

$$s = -\frac{\epsilon + \kappa}{2} + \sqrt{\left(\frac{\epsilon + \kappa}{2}\right)^2 + s\;\epsilon} \tag{A8}$$

Assuming that the spread of the variant is slow relative to the latent and infectious periods ($s \ll \epsilon, \kappa$), adding a latent period causes selection to weaken slightly, with *Equation A8* approaching $s\frac{1}{1+\,\kappa/\epsilon}$. This occurs because only during a fraction $\frac{1/\kappa}{1/\epsilon+1/\kappa} = \frac{1}{1+\,\kappa/\epsilon}$ of the generation time of the virus is it infectious. Xin et al. (2023) estimate a mean latent period of 3.1 days for Omicron. For the parameters considered typical of Omicron (Appendix A), selection would be ~70% as strong with a latent period. We ignore this correction to simplify the model presentation.

Seroconversion

Another real-world complication is that not all individuals seroconvert following infection or vaccination (i.e., not all infections elicit a robust immune response). If a fraction q of infections boost immunity (meaning here that they recover to the R_1 compartment in the SIR_n model), while $1–q$ become susceptible again (returning to the S compartment), the equilibrium fraction of infected individuals changes to:

$$\hat{I} = \left(1 - \frac{\kappa}{\beta}\right)\frac{\delta/q}{\delta/q + \kappa}. \tag{A9a}$$

Thus, decreasing the seroconversion rate by a factor q has the same effect on the endemic equilibrium as increasing the waning rate by a factor $1/q$ (*Equation 5*), and the same holds for the selection coefficient of a variant, s (*Equation 4*, Supplementary *Mathematica* package). The temporal dynamics of the wavelets are slightly different, with immediate waning for those who do not seroconvert and slower waning for those who do (Supplementary *Mathematica* package).

Seroconversion rates for vaccines can also be included, changing the equilibrium to:

$$\hat{I} = \left(1 - \frac{\kappa}{\beta}\right)\frac{\delta/q}{\delta/q + \kappa} - \frac{\upsilon\;q_\upsilon/q}{\delta/q + \kappa}. \tag{A9b}$$

where q_υ is the seroconversion rate for vaccination (here meaning the probability that a vaccine dose boosts antibodies and provides protection from infection). If seroconversion rates are similar following infection and vaccination ($q_\upsilon = q$), then the results for selection (*Equation 3*) and endemic incidence (*Equation 4*) are again the same if we replace δ (ignoring seroconversion) with δ/q (including it).

To simplify the presentation, we do not explicitly include seroconversion but consider a range of waning rates to cover both seroconversion and waning.

Empirically, high seroconversion rates have been reported following vaccination with a single dose of Pfizer's BNT162b2 (q = 99.5%) or AstraZeneca ChAdOx1 (q = 97.1%), leading to antibodies recognizing the spike protein (Wei et al., 2021). Slightly lower seroconversion (q = 93.5%–95.3%) was observed following infection in early 2020 (Oved et al., 2020). An estimate following Omicron infection inferred even lower rates of seroconversion of q = 74%–81% (here examining antibodies to nucleocapsid, as anti-spike antibodies were nearly universal in the highly vaccinated population examined; Erikstrup et al., 2022).

Multiple variants

To determine the robustness of our results to the assumption that the population is at equilibrium, we introduce two variants into the population separated by a time interval of varying length. We then ask whether the model predictions for the second variant are substantively altered by the spread of the first variant. We explore the case where the first variant is more transmissible and the second more immune evasive (Supplementary Figure S4, top row) and vice versa (Supplementary Figure S4, bottom row).

Alternative vaccination models

In our core model, we considered a vaccination campaign aimed at vaccinating a given number of susceptible individuals every day. Here we briefly summarize results for two alternative assumptions that were modeled in the Supplementary *Mathematica* package:

- If susceptibilities are vaccinated at a per capita rate (replacing $-\nu$ in *Equation 1* with $-\nu_S S$), the equilibrium is unaffected given the same daily rate of vaccinations at the endemic equilibrium (i.e., setting $\nu_S = \nu/\hat{S}$), but the numerical dynamics are slightly different because the number of vaccines per day fluctuates as the number of susceptible individuals fluctuates.
- If vaccines are distributed to all individuals, regardless of past infection status, then vaccinations are less effective at reducing the incidence of disease. In this alternate model, we assumed that all individuals were vaccinated at per capita rate ν, moving from their previous disease class into the first recovered class, R_1. Compared to Figure 6A, vaccinating all individuals caused the equilibrium to fall by only 23.6%, rather than 41.8%, because of the inefficiency of vaccinating individuals who are already immune.

Appendix C: Non-pharmaceutical interventions

We consider an expansion of the SIR$_n$ epidemiological model to allow heterogeneity in behavior. As illustrated in Supplementary Figure S5, we now allow two classes of individuals, those who regularly adhere to stronger NPI measures, such as masking (indicated by an ‡), and those who do not:

$$\begin{aligned}
\frac{dS}{dt} &= \delta_n R_n - \beta SI - (1-p)\,\beta SI^{\ddagger} \\
\frac{dS^{\ddagger}}{dt} &= \delta_n R_n^{\ddagger} - (1-p)\,\beta S^{\ddagger} I - (1-p)^2 \beta S^{\ddagger} I^{\ddagger} \\
\frac{dI}{dt} &= \beta SI + (1-p)\,\beta SI^{\ddagger} - \kappa I \\
\frac{dI^{\ddagger}}{dt} &= (1-p)\,\beta S^{\ddagger} I + (1-p)^2 \beta S^{\ddagger} I^{\ddagger} - \kappa\, I^{\ddagger} \\
\frac{dR_1}{dt} &= \kappa\, I - \delta_1 R_1 \\
\frac{dR_1^{\ddagger}}{dt} &= \kappa\, I^{\ddagger} - \delta_1 R_1^{\ddagger} \\
\frac{dR_j}{dt} &= \delta_{j-1} R_{j-1} - \delta_j R_j \\
\frac{dR_j^{\ddagger}}{dt} &= \delta_{j-1} R_{j-1}^{\ddagger} - \delta_j R_j^{\ddagger} \qquad \text{for } 2 \le j \le n
\end{aligned} \tag{A10}$$

where the last two equations are repeated for the remaining waning classes (j from 2 to n). We again set all rates between waning classes to $\delta_i = \delta/n$ (mean waning time of $1/\delta$ days). The new parameter p measures the protection provided when one individual in an interaction engages in NPI measures (reducing β by a factor 1–p). If both infected and susceptible individuals uphold these measures, transmission is reduced by $(1-p)^2$. All variables are measured as proportions of the total population, with f being the fraction of the population carrying out NPI measures, such as masking (the sum of the ‡ variables).

There are two equilibria of this system *Equation A10*, one where the disease is absent and one where the disease is endemic at:

$$\begin{aligned}
\hat{S} &= \frac{1}{2}\left(-B + \sqrt{B^2 - 4C}\right) \\
\hat{S}^{\ddagger} &= \frac{\kappa - \beta\,\hat{S}}{(1-p)^2 \beta} \\
\hat{I} &= \left(1 - f - \hat{S}\right)\frac{\delta}{\delta + \kappa} \\
\hat{I}^{\ddagger} &= \left(1 - f - \hat{S}\right)\frac{\delta}{\delta + \kappa}\,\frac{\kappa - \beta\,\hat{S}}{(1-p)\;\beta\,\hat{S}} \\
\widehat{R}_j &= \widehat{R}_{j-1} = \frac{1 - f - \widehat{S} - \widehat{I}}{n} \qquad \text{for } 1 \le j \le n \\
\widehat{R}_j^{\ddagger} &= \widehat{R}_{j-1}^{\ddagger} = \frac{f - \widehat{S} - \widehat{I}}{n} \qquad \text{for } 1 \le j \le n
\end{aligned} \tag{A11}$$

where $B = \frac{(1-f)\;(1-p)\;\beta + f\;(1-p)^2\beta - \kappa\;p}{p\;\beta}$ and $C = -\frac{(1-f)\;(1-p)\;\kappa}{p\;\beta}$. The disease-absent equilibrium is locally stable when transmission rates are low relative to clearance, such that the endemic reproductive number if everyone were susceptible is less than one, $\tilde{R}_0 = \frac{(1-f)\;\beta + f\;(1-p)^2\beta}{\kappa} < 1$, in which case the endemic equilibrium does not exist (i.e., not all variables are positive). Otherwise, when $\tilde{R}_0 > 1$, the endemic equilibrium exists and is stable for the parameters considered, but it may become unstable for large n (Hethcote et al., 1981).

At this equilibrium, the risk that an individual is in the infected class at any point in time is $\hat{I}^{\ddagger}/f$ if they regularly mask and $\hat{I}/(1-f)$ if they do not, from which we calculate the relative risk in the main text. The population-level impact of NPI measures, such as masking, is determined by analyzing the fraction of the population expected to be infected at any point in time, $\hat{I} + \hat{I}^{\ddagger}$.

The above assumes that an individual's choice about engaging in NPI measures remains constant over time, but we also consider the opposite case (detailed in the Supplementary *Mathematica* package), where individuals rapidly switch between engaging or not in NPI measures. Assuming that the behavior persists over the short time frame of infection but that individuals switch often while in the longer susceptible or recovered phases, we can simplify the model by monitoring only those engaging in NPI measures at the time of exposure, with f then representing the probability that an individual engages in the NPI measures at that time. We thus only subdivide the infectious class into those who were or were not practicing the NPI measures at the time of infection ($I^{\ddagger}$ or I, respectively). The dynamics are then:

$$\begin{aligned}
\frac{dS}{dt} = {} & \delta_n R_n - (1-f)\,\beta SI \\
& - (1-p)\,f\beta SI - (1-p)\,(1-f)\,\beta SI^{\ddagger} \\
& - (1-p)^2 f \beta SI^{\ddagger} \\
\frac{dI}{dt} &= (1-f)\,\beta SI + (1-p)\,(1-f)\,\beta SI^{\ddagger} - \kappa\, I \\
\frac{dI^{\ddagger}}{dt} &= (1-p)\,f\beta SI + (1-p)^2 f \beta SI^{\ddagger} - \kappa\, I^{\ddagger}
\end{aligned} \tag{A12}$$

$$\frac{dR_1}{dt} = \kappa\, I + \kappa\, I^{\ddagger} - \delta_1 R_1$$

$$\frac{dR_j}{dt} = \delta_{j-1} R_{j-1} - \delta_j R_j \qquad \text{for } 2 \le j \le n$$

Results using *Equation A12* instead of *Equation A10* are similar, except that practicing and non-practicing individuals are equally likely to be susceptible at the time of exposure, so the individual-level protective effect of the NPI measure now depends only on p and not on $\tilde{R}_0$, as discussed in the text.

Parameters: The protection (p) and uptake (f) depend on the NPI measure considered (The Royal Society, 2023). Here we briefly review data on masking as a protective measure. One meta-analysis of randomized control studies prior to the COVID-19 pandemic found protection provided by masks was $p = 16\%$ for respiratory infections, rising to $p = 24\%$ in studies longer than two weeks (Li et al., 2022). Importantly, many individual studies were underpowered, but the results were consistent across studies (see Figure 2 in Li et al., 2022).

For COVID-19, a meta-analysis of the impact of mask mandates estimated a 25% reduction in transmission rates, comparing transmission levels predicted if everyone were in the class that self-report wearing masks "most of the time in some public places" to that if no one wore masks (Leech et al., 2022). Importantly, the authors showed that the lifting or imposition of mandates rarely had dramatic immediate effects on mask-wearing, emphasizing that mandates are a poor proxy for mask-wearing. Their analysis thus benefited from a global analysis of trends in mask-wearing behavior around the time of mandates by incorporating data from a survey of masking behavior among nearly 20 million individuals.

As argued by Leech et al. (2022), this effect size is likely to be underestimated for a number of reasons. First, the study period (May 1–September 1, 2020) occurred when cloth masks predominated, because high-quality masks were largely unavailable outside of healthcare settings. Second, the definition of mask use was broad and included individuals who only occasionally mask and do so in a few public places. We thus consider that $p = 0.25$ represents a lower bound on the protection provided by masking.

Higher values are plausible when using high-quality masks and doing so consistently in indoor public spaces. For example, masks provided a stronger benefit, reducing the odds ratio of infection by an average of 50% among the studies summarized within healthcare settings (The Royal Society, 2023). We thus consider $p = 0.5$ to represent a reasonable upper bound on the protection provided by masking attainable by consistent wearing of high-quality masks. Combinations of NPI measures, including improved ventilation, avoiding crowded indoor environments, testing and self-isolation, and masking may provide considerably stronger protection.