Face masking and COVID-19: potential effects of variolation on transmission dynamics

Zachary Levine and David J. D. Earn

Department of Mathematics and Statistics, McMaster University, Hamilton, Canada L8S 4K1

Cite this article: Levine Z, Earn DJD. 2022
Face masking and COVID-19: potential effects of variolation on transmission dynamics.
J. R. Soc. Interface 19: 20210781.
https://doi.org/10.1098/rsif.2021.0781

Received: 8 October 2021
Accepted: 4 April 2022

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biomathematics

Keywords:
infectious diseases, COVID-19, variolation, face masks, SIR model, differential equations

Author for correspondence:
Zachary Levine
e-mail: zachary.levine@weizmann.ac.il

Face masks do not completely prevent transmission of respiratory infections, but masked individuals are likely to inhale fewer infectious particles. If smaller infectious doses tend to yield milder infections, yet ultimately induce similar levels of immunity, then masking could reduce the prevalence of severe disease even if the total number of infections is unaffected. It has been suggested that this effect of masking is analogous to the pre-vaccination practice of variolation for smallpox, whereby susceptible individuals were intentionally infected with small doses of live virus (and often acquired immunity without severe disease). We present a simple epidemiological model in which mask-induced variolation causes milder infections, potentially with lower transmission rate and/or different duration. We derive relationships between the effectiveness of mask-induced variolation and important epidemiological metrics (the basic reproduction number and initial epidemic growth rate, and the peak prevalence, attack rate and equilibrium prevalence of severe infections). We illustrate our results using parameter estimates for the original SARS-CoV-2 wild-type virus, as well as the Alpha, Delta and Omicron variants. Our results suggest that if variolation is a genuine side-effect of masking, then the importance of face masks as a tool for reducing healthcare burdens from COVID-19 may be under-appreciated.

1. Introduction

Early in the COVID-19 pandemic, face masking was discouraged outside healthcare settings [1,2], and inadequate supplies of masks created significant challenges for healthcare workers [3–5]. By 3 April 2020 appreciation of the potential value of masking had increased sufficiently for the US Centers for Disease Control and Prevention (CDC) to recommend the wearing of face coverings in public [6]. Similar recommendations were made in Canada on 6 April 2020 [7] and in the UK on 11 May 2020 [8]. Over the course of the pandemic, evidence that masking is an effective tool to reduce community transmission of SARS-CoV-2 has continued to accumulate [9–16].

An additional potential benefit of masking—even in situations where it fails to prevent transmission—was proposed in the summer of 2020 by Gandhi & Rutherford [17] and Gandhi et al. [18]. They noted that if transmission occurs in the presence of face masks, then the infecting viral inoculum is likely to be smaller than is typical when masks are not worn. If smaller inocula tend to lead to less severe infections, then—even if masks fail to block transmission completely—masking could reduce morbidity and mortality from COVID-19 and boost the level of herd immunity in the population.

The notion that promoting mild infections could be an effective disease control strategy has a long history. In the eighteenth century, it was common to infect children with smallpox intentionally—a process known as variolation—by administering a small inoculum of smallpox virus taken from an infected
person [19–24]. While SARS-CoV-2 infections are never intentional, Gandhi & Rutherford [17] refer to SARS-CoV-2 transmission via small ino
culations that penetrate masks as variola
tion. In this paper, we explore the potential benefits of SARS-CoV-2 variolation using a mathematical model.

We begin in §2 by reviewing evidence that supports the hypothesis that face masking may promote SARS-CoV-2 variolation. In subsequent sections, we present and analyse a simple model that allows us to investigate potential effects of variolation induced by face masking.

2. Variolation as a side effect of face masking

Face masks can reduce the probability of transmission either through outward or inward filtration. Outward filtration occurs when droplets containing viral ino
culations are captured when leaving the mouth or nose of an infected person [25]. Inward filtration occurs when viral particles are prevented from entering someone’s nose or mouth.

Gandhi & Rutherford [17] suggested that people who are infected with SARS-CoV-2 when wearing a face mask might experience less severe illness, because the size of the viral ino
culations they receive is reduced by inward filtration. Assuming that individuals infected in this way still develop lasting immunity, they can be considered variolated. Variolation in this sense might not reduce the total proportion of the population infected, but would nevertheless benefit the population by increasing the proportion of infections that are mild or asymptomatic [17].

The variolation hypothesis, as originally formulated by Gandhi & Rutherford [17], depends on three key assumptions:

**Reduced ino
culations:** individuals who are infected while wearing a mask receive a smaller viral ino
culation than if they had not been masked.

**Reduced severity:** smaller viral ino
culations tend to yield less severe infections with milder symptoms, i.e. there is a positive dose–response relationship.

**Acquired immunity:** mild infections still provide long-lasting natural immunity to the disease.

Reduced ino
culations is plausible because masks are known to filter droplets that may contain virus particles [26].

Reduced severity from smaller ino
culations is indicated by several lines of evidence. The dose–response relationship is supported by an investigation in Madrid that found that ‘distinct sizes of viral ino
culations at the time of exposure’ could explain different illness courses in three clusters of SARS-CoV-2 infection (infecteds in different clusters developed COVID-19 with different severity, and the only clear difference among the clusters was the degree of SARS-CoV-2 exposure) [27]. In addition, disproportionately severe disease outcomes were documented following a choir rehearsal in the state of Washington in March 2020 [28]; the observed severity of illness among attendees is consistent with a positive dose–response relationship because infected individuals likely spew more virus particles when singing, compared to simply talking at a social gathering. Finally, an experimental study exposed hamsters to SARS-CoV-2 in a laboratory and found that COVID-19 symptoms were much less severe in infected animals that were shielded by a surgical mask partition [29].

Immunity after mild infection is suggested by a 2020 study in the New York City area, where 13.7% of a sample of healthcare personnel (HCP) were found to have antibodies to SARS-CoV-2 [30]; HCP were not asked if they were symptomatic, but 5% (11%) of those who tested positive for SARS-CoV-2 antibodies reported low (medium) likelihood of SARS-CoV-2 exposure. Another study found that 15% of those with SARS-CoV-2 antibodies reported never having COVID-19 symptoms [31]. Other investigations have revealed that antibodies to SARS-CoV-2 were present in individuals five months after asymptomatic or mild infections [32,33]. Thus, asymptomatic and mild SARS-CoV-2 infections likely do provide at least some protection against reinfection.

3. Model

We investigate the potential effects of facemask-induced variolation by expanding the standard Susceptible–Infectious–Removed (SIR) model [34,35].

We assume that all members of the population are identi
cal—and that the population is homogeneously mixed—so we do not distinguish between situations in which poor face
caps are worn universally by everyone versus situations in which better masks are worn by a subset of the population; thus, masks are implicitly assumed to be distributed to individuals at random.

We capture the variolating effect of facemasks by imagin
ing that adherence to facemasking causes a proportion $m$ of infections to be mild, and we therefore refer to $m$ as the *probability of mild infection*. Mild infections might be shorter and/or less transmissible (however, the *effective*) infectious period for severe infections could, in practice, be shorter than for mild infections, because severe cases are likely to be isolated quickly. Thus, three distinct effects—causing mild infections, reducing transmission rate and potentially shortening infectious periods—contribute to the overall *effectiveness of mask-induced variolation*. We allow for the possi
bility that immunity decays [36], but assume the duration of immunity is the same following mild or severe infections.

Figure 1 presents a flow chart for the model, which we represent formally as a system of ordinary differential equations (ODEs):

\[
\frac{dS}{dt} = -\beta S I - \gamma S + \delta N - \mu S - \mu R - \delta R,
\]

\[
\frac{dI_m}{dt} = \mu S - \gamma I - \mu I_m + \mu I_m,
\]

\[
\frac{dI_l}{dt} = (1 - m) S I - \gamma I - \mu I_l + \mu I_l,
\]

\[
\frac{dR_m}{dt} = \gamma I - \mu R_m - \delta R_m + \mu R_m,
\]

\[
\frac{dR_l}{dt} = \gamma I - \mu R_l - \delta R_l + \mu R_l.
\]

where the force of infection is

\[
\Lambda = \beta m \frac{I_m}{N} + \beta s \frac{I_s}{N}.
\]

and the total population size is

\[
N = S + I_m + I_l + R_m + R_l.
\]

All the state variables are listed in table 1. $R_m$ and $R_l$ do not appear in the first three lines of equation (3.1) and can be ignored for the purposes of dynamical analysis. We retain equations (3.1d,e) for convenience, to keep track
of the numbers of immune individuals who had mild versus severe infections.

Estimated values of parameters associated with COVID-19 are listed in table 2. We use these estimates to set default values for the parameters of our model, which are listed in table 3. Since we ignore the latent stage in our model, we use the estimated mean generation interval for COVID-19 in place of the mean infectious period in our model (cf. [49,52]). We ignore disease-induced mortality, but we (more than) compensate for this by assuming for simplicity that the death rate from other causes ($\mu_s$) is equal to the birth rate ($\nu$); thus, births are balanced by deaths and the population size ($N$) is constant. Figure 2 shows prevalence and cumulative incidence time series, obtained by solving equation (3.1) numerically with the default parameters.

4. Analysis

4.1. Basic reproduction number ($R_0$)

The contribution to $R_0$ from infectors who are mildly infectious is the probability of mild infection ($m$), times the transmission rate from mildly infectious individuals ($b_m$), times the expected time that a mildly infectious individual is infectious ($1/(\gamma_m + \mu)$). There is a similar contribution from severely infectious individuals, hence

$$R_0 = m \frac{\beta_m}{\gamma_m + \mu} + (1 - m) \frac{\beta_s}{\gamma_s + \mu}. \quad (4.1)$$

A formal calculation, applying the method of [54] to equation (3.1), yields the same expression.

4.2. Dimensionless stage durations

The mean time spent in the mildly infectious state, as a fraction of the mean lifetime, is

$$e_m = \frac{\mu}{\gamma_m + \mu}. \quad (4.2a)$$

Similarly, for severe infections, we have

$$e_s = \frac{\mu}{\gamma_s + \mu}. \quad (4.2b)$$

The mean across both types of infections is

$$e = m e_m + (1 - m) e_s. \quad (4.2c)$$

It is worth noting that

$$e \leq \max \{e_m, e_s\} < 1. \quad (4.3)$$

The mean duration of immunity, as a fraction of the mean lifetime, is

$$\eta = \frac{\mu}{\delta}. \quad (4.4)$$
Table 2. Parameter estimates (and 95% confidence intervals) for COVID-19; these estimates are examples from published studies and are not based on a systematic review of the literature. WT refers to the wild-type virus. The latent period (time from exposure to infectiousness) is extremely difficult to estimate, so the incubation period (time from exposure to appearance of symptoms) is often used as a proxy. The incubation period for the Alpha variant was estimated to be $\frac{1}{2} \times (T_{inc} \text{ for WT})$ [37]. $R_0$ is expressed in terms of the model parameters in equation (4.1). $R_0$ for the Alpha, Delta and Omicron variants were approximated using estimates of the increase in transmissibility for each successive variant, i.e. WT $\frac{1.5^{\Delta}}{2} \times \Delta$ Alpha [38], Alpha $\frac{1.5^{\Delta}}{2} \times \Delta$ Delta [39–41] and Delta $\frac{4.2 \times [2.1 \times \Delta]}{38}$ Omicron [42].

| parameter | meaning | estimate | references |
|-----------|---------|----------|------------|
| $T_{lat}$ | mean latent period | 3.7 [3.3–3.9] days | WT [43] |
| $T_{inc}$ | mean incubation period | 4.0 [3.5–4.4] days | Delta [44] |
| $T_{inf,m}$ | mean infectious period for mild infections | 6.4 [4.9–8.5] days | WT [45] |
| $T_{inf,s}$ | mean infectious period for severe infections | 4.2 [3.2–5.6] days | Alpha [37] |
| $T_{gen,m}$ | mean generation interval for mild infections | 5.8 [5.2–6.4] days | Delta [44] |
| $T_{gen,s}$ | mean generation interval for severe infections | $\sim 3$ days | Omicron [46] |
| $T_{imm}$ | mean immune period | 9 [6–11] days | [47] |
| $R_0$ | basic reproduction number | 14 [8–20] days | [48] |
| $T_{imm,m}$ | mean incubation period for mild infections | $T_{lat} + T_{inf,m}$ days | [49, equation (4.1)] |
| $T_{imm,s}$ | mean incubation period for severe infections | $T_{lat} + T_{inf,s}$ days | [49, equation (4.1)] |
| $R_0$ | basic reproduction number | 7 [6–8] months | [32,33,50] |
| $T_{lat}$ | mean latent period | 3 [2.1–4.6] | WT [38] |
| $T_{inc}$ | mean incubation period | 4.5 [3.15–6.9] | Alpha [51] |
| $T_{inf,m}$ | mean infectious period for mild infections | 6.75 [4.7–10.4] | Delta [39–41] |
| $T_{inf,s}$ | mean infectious period for severe infections | 28.4 [9.9–94.6] | Omicron [42] |

4.3. Equilibria and stability
The system described by equation (3.1) always has a disease free equilibrium (DFE), at which $S = N$ and all other compartments are empty. In addition, if $R_0 > 1$ then there is an endemic equilibrium (EE) given by

$$
\frac{S}{N} = \frac{1}{R_0},
$$

$$
\frac{I_m}{N} = m e_m \left(1 - \frac{1}{R_0}\right) \frac{\eta + 1}{\eta + e},
$$

and

$$
\frac{I_s}{N} = (1 - m) e_s \left(1 - \frac{1}{R_0}\right) \frac{\eta + 1}{\eta + e}.
$$

In the limit of permanent immunity ($\delta \to 0$, $\eta \to \infty$), the final factors in equations (4.5b,c) simplify to 1. Since $e < 1$, decay of immunity ($\eta < \infty$) necessarily increases equilibrium prevalence. Theorem 2 of [54] establishes that $R_0 = 1$ is the boundary between local stability and instability of the DFE. In fact, the model (3.1) is a special case of the class of SIR models with multiple parallel infectious stages considered by Korobeinikov [55]. Consequently, theorem 1 of [55] establishes that the DFE is globally asymptotically stable (GAS) if $R_0 \leq 1$, and that the EE is GAS if $R_0 > 1$.

4.4. Initial growth rate ($r$)
The initial exponential growth rate of an epidemic beginning near the DFE of equation (3.1) is the largest eigenvalue of the Jacobian derivative of the vector field $(S, I_m, I_s)$ at the DFE. This Jacobian has a first column $(-\delta - \mu, 0, 0)$, so $-(\delta + \mu)$ is an eigenvalue. The other two eigenvalues are determined by the submatrix

$$
\begin{pmatrix}
mb_m - (\gamma_m + \mu) & mb_s \\
(1 - m)\beta_m & (1 - m)\beta_s - (\gamma_s + \mu)
\end{pmatrix}
$$

(4.6)

If we write, for convenience,

$$
\beta = m\beta_m + (1 - m)\beta_s,
$$

then the larger of the two eigenvalues of (4.6) is

$$
r = \frac{1}{2} \left( \beta - (\gamma_m + \gamma_s + 2\mu) + \sqrt{\left(\beta - (\gamma_m - \gamma_s)\right)^2 - 4mb_m(\gamma_m - \gamma_s)} \right).
$$

(4.8)

Note that, as written, the discriminant in equation (4.8) is manifestly positive if $\gamma_m \leq \gamma_s$, but it can also be written

$$
\left(\beta - (\gamma_m - \gamma_s)\right)^2 + 4(1 - m)\beta_s(\gamma_m - \gamma_s),
$$

(4.9)

which is manifestly positive if $\gamma_m \geq \gamma_s$. In the limit of equal mild and severe infectious periods, the initial growth rate is simply

$$
r = \beta - (\gamma + \mu), \text{ if } \gamma_m = \gamma_s = \gamma.
$$

(4.10)

For any initial growth rate $r$ (equation (4.8)), the initial doubling time, i.e. the time required for prevalence to double during the exponential growth phase, is

$$
T_2 = \frac{\log 2}{r}.
$$

(4.11)
4.5. Final size

If there is no source of new susceptibles ($\nu = \mu = \delta = 0$), we can derive a final size relation based on equation (3.1). Following [56,57], we look for a constant of the motion that is the sum of log($S/N$) and a linear combination of the other state variables. Solving for coefficients that cause the time-derivative of this expression to vanish, we find that

$$F(t) = \log \frac{S}{N} - (1 - m) \left( \frac{\beta_m}{\gamma_m} - \frac{\beta_s}{\gamma_s} \right) \frac{I_m}{N} + m \left( \frac{\beta_m}{\gamma_m} - \frac{\beta_s}{\gamma_s} \right) \frac{I_s}{N}$$

$$+ \left( m \frac{\beta_m}{\gamma_m} + (1 - m) \frac{\beta_s}{\gamma_s} \right) \frac{R_m + R_s}{N}$$

is a constant of the motion. In particular, $F(0) = F(\infty)$; consequently, since $(S, I_m, I_s) \to (N, 0, 0)$ as $t \to 0$ and $(I_m, I_s) \to (0, 0)$ as $t \to \infty$ (and noting that the coefficient of the final term in equation (4.12) is $R_0$) we have

$$0 = \log \frac{S(\infty)}{N} + R_0 \left( \frac{R_m(\infty) + R_s(\infty)}{N} \right).$$

(4.13)

Writing $Z = 1 - S(\infty)/N$ and noting that

$$\frac{R_m(\infty) + R_s(\infty)}{N} = Z_s,$$

we obtain the standard final size relation [34,57],

$$\log (1 - Z) = R_0 Z.$$  

(4.15)

With the same approach, but insisting that the coefficient of $R_s$ vanishes, we find another constant of the motion

$$F_m(t) = \log \frac{S}{N} + \frac{1 - m \beta_s I_m}{\gamma_s N} - \frac{\beta_m I_m}{\gamma_m N} + \frac{1}{m} R_0 \frac{R_m}{N}.$$

(4.16)

Considering the limits $t \to 0$ and $t \to \infty$ yields

$$\frac{R_m(\infty)}{N} = m R_0 \log (1 - Z).$$

(4.17)

From this expression and equations (4.14) and (4.15), we therefore have exact expressions for the mild and severe attack rates,

$$Z_m = \frac{R_m(\infty)}{N} = m Z$$

and

$$Z_s = \frac{R_s(\infty)}{N} = (1 - m) Z.$$  

(4.18)

Solving equation (4.15) for $Z$ [57], we can write

$$Z_s(m) = (1 - m) \left( 1 + \frac{1}{R_0(m)} W_0 \left[ - R_0(m) e^{-R_0(m)} \right] \right),$$

(4.19)

where $W_0$ is the principal branch of Lambert's $W$ function [58]. In equation (4.19), we have emphasized the dependence of $Z_s$ on $m$ (including the dependence of $R_0$ on $m$, equation (4.1)), since we are especially interested in understanding how an increase in the proportion of cases that are mild influences the expected number of severe infections.
Table 3. Parameters of the model described in figure 1 and equation (3.1). The probability with which masking causes infections to be mild (m) is unknown. Our default value is chosen to be substantial so that for illustrative graphs constructed with fixed m (figures 5, 6 in §5) the effect of mask-induced variation is non-negligible. The recovery rates can be interpreted as the rates of ‘recovery or death’ since we do not explicitly model disease-induced mortality (cf. final paragraph of §3). The death rate μ refers to mortality from causes other than the focal disease. Note that mild illness is assumed to be associated with mild infectiousness. All birth rates were estimated for the years 2015–2020. We use a default latent period of $T_{\text{lat}} = 3.7$ days for all variants. The generation interval for an SEIR model is $T_{\text{gen}} = T_{\text{lat}}$ (49, eqn (4.1)). Setting $\gamma_s = 1/T_{\text{gen},\text{s}}$ in our model yields dynamics more similar to an SEIR version (cf. (49,52)), so it is a better approximation of the real world than an SIR version with $1/\gamma$ taken to be the observed mean infectious period. The transmission rate for severe infections ($\beta_s$) is set for each variant using equation (4.1) with $m = 0$ and the associated $R_s$ estimate listed in table 2. After specifying $\beta_s$, we then set $\beta_m = (\beta_s/\beta_s) \times \beta_s$.

| parameter | meaning | expression or default value |
|-----------|---------|----------------------------|
| $m$       | probability an infected individual develops mild illness | 0.6 |
| $\beta_m$ | transmission rate from mildly infectious individuals | equation (4.1) |
| $\beta_s$ | transmission rate from severely infectious individuals | equation (4.1) |
| $\beta_m/\beta_s$ | ratio of transmission rates | 1/2 |
| $\gamma_m$ | recovery rate from mild infections | $1/T_{\text{gen},\text{m}}$ |
| $\gamma_s$ | recovery rate from severe infections | $1/T_{\text{gen},\text{s}}$ |
| $\nu$ | per capita annual birth rate [53] | 0.0105 (Canada) |
| $\delta$ | per capita annual death rate | 0.0115 (UK) |
| $\mu$ | rate of decay of immunity | 0.012 (USA) |

4.6. Peak prevalence

Another way of writing equation (4.12) is

$$\frac{1}{R_0} \left( \frac{\beta_m I_m}{\gamma_m N} + \frac{\beta_s I_s}{\gamma_s N} \right) = \left( 1 - \frac{S}{N} \right) + \frac{1}{R_0} \log \frac{S}{N} \quad (4.20)$$

which does not depend explicitly on the probability $m$ that an infection is mild (there is implicit dependence on $m$ through $R_0$, equation (4.1)). In the limit that mild and severe infections have the same reproduction number ($\beta_m/\gamma_m = \beta_s/\gamma_s$), the left-hand side of equation (4.20) reduces to the total prevalence,

$$\frac{I}{N} = \frac{I_m + I_s}{N} \quad (4.21)$$

and equation (4.20) agrees exactly with the formula for the phase portrait $I(S)$ of the standard SIR model. The right-hand side of equation (4.20) is maximized at $S = N/R_0$, so inserting $S = N/R_0$ into the equation (4.20) yields the peak prevalence formula for the standard SIR model,

$$\frac{I_{\text{peak}}}{N} = 1 - \frac{1}{R_0} \left( 1 + \log R_0 \right) \quad (4.22)$$

This formula will provide a good approximation to the peak total prevalence in our model (equation (3.1)) to the extent that the left-hand side of equation (4.20) approximates total prevalence near its peak. Figure 3 compares total prevalence (equation (4.21)) with the left-hand side of equation (4.20) for the full range of possible mask-induced increase in the probability of mild infection ($0 \leq m \leq 1$) with other parameters fixed at the values specified in tables 2 and 3. The figure indicates that approximating peak prevalence with the left-hand side of equation (4.20) is reasonable (the maximum relative error is 5.5%).

If we approximate the susceptible population $S$ at peak total prevalence by $N/R_0$, then inserting $S = N/R_0$ into equation (4.20) provides an estimate of how prevalence at its peak is partitioned between mild and severe cases. Of course, we can approximate the value of $S$ at peak total prevalence to any desired accuracy by solving equation (3.1) numerically; with that approach, we can use equation (4.20) to determine, to any desired accuracy, the partitioning of peak prevalence between mild and severe cases.

In practice, the peak prevalence of severe cases is what is most important from the point of view of stress on health care systems, and it is simplest to find the peak of $I_s$ directly from numerical solutions of equation (3.1). However, it is important to bear in mind that $I_s$ is the number of individuals who are still contributing to transmission dynamics but are suffering from disease that is so severe that they will need substantial healthcare. $I_s$ does not include people who have already been isolated in hospitals or other settings (and can therefore be considered unable to cause further infections). Thus, $I_s$ is a measure of upcoming healthcare demand, as opposed to the current burden on the system. The peak healthcare burden can be estimated roughly by multiplying the peak of $I_s$ by $T_{\text{H}}/T_{\text{gen},\text{s}}$, where $T_{\text{H}}$ is the mean length of stay in hospital and $T_{\text{gen},\text{s}} = 1/\gamma_s$ is the mean time spent in the $I_s$ compartment of our model (cf. tables 2 and 3).

5. Illustration of results for COVID-19

Figure 4 shows how several important epidemiological risk metrics depend on the variating effect of masking, if it acts principally by increasing the probability of mild infection ($m$). Each panel shows three curves, associated with the estimated basic reproduction number for the original wild-type SARS-CoV-2 virus (WT, $R_0 \sim 3$; [38]), the Alpha variant that began to spread in late 2020 ($R_0 \sim 4.5$; [51]), the Delta variant that emerged in the spring of 2021 ($R_0 \sim 6.75$;
and the Omicron variant that emerged in late 2021 ($R_0 \approx 28$; [42]). $R_0$ for each variant is assumed to take the observed value in the limit of no variolating effect ($m = 0$) and then to decrease according to equation (4.1). All other parameters are fixed at the default values listed in tables 2 and 3; in particular, we assume that the transmission rates of mild and severe infections ($\beta_m, \beta_s$) are independent of their probability of occurrence ($m, 1 - m$).

Figure 4b shows that the initial growth rate of the epidemic is strongly dependent on the probability $m$ and is more sensitive to $m$ if the variant is more transmissible. As $m$ is increased from 0 to 1, the initial doubling time increases from 5.1 to 65 days for WT (upper limit not shown on the graph), 2.9 to 11 days for Alpha, 1.8 to 5.0 days for Delta and 0.37 to 0.79 days for Omicron. (Note that these are estimates of expected doubling times in a completely susceptible population, whereas the doubling times actually observed when the later variants emerged were in populations that already had substantial levels of immunity from previous infections and vaccination.)

Figure 4c–e shows that risk measures related to severe infections decline substantially with $m$, as expected since severe illness is completely eliminated in the limit of perfect variolation ($m \to 1$; all cases mild). Variolation has relatively greater effect (on prevalence of severe illness) for more transmissible variants.

Figures 5 and 6 show how the same risk metrics depend on the relative transmissibility of mild infections ($\beta_m / \beta_s$) and the relative length of mild infections ($\gamma_m / \gamma_s$). The horizontal scale in figure 5 ends at 1 because it is implausible that mild infections are intrinsically more transmissible than severe infections (e.g. severe cases are likely to be isolated sooner and more stringently). The similarities between figures 5 and 6 can be attributed to the fact that increasing $\beta_m$ or decreasing $\gamma_m$ has a similar effect on $R_0$ (equation (4.1)).

6. Discussion

We have explored the potential role of mask-induced variolation in reducing the impact of COVID-19 (or other directly transmitted infectious diseases). Our approach has been to analyse a highly idealized mathematical model (figure 1
and equation (3.1)) that captures the key mechanisms that we wished to investigate, namely the potential for masking to cause a proportion of infections to be mild, reduce the probability of infection upon contact and/or change the infectious period. Because our model does not attempt to be realistic in detail, conclusions we draw are qualitative only. However, the simplicity of the model has made it possible to derive results analytically, so qualitative conclusions
can easily be drawn for pathogens with reproduction numbers or generation intervals that are different from those of SARS-CoV-2.

6.1. Summary of results

Our main results are analytical formulae that show how the model parameters—including the probability of mask-
induced mild infection ($m$), the relative transmissibility of mild infections ($\beta_m/\beta_s$) and the relative length of mild infections ($\gamma_s/\gamma_m$)—influence the initial epidemic doubling time (equation (4.11)), the peak prevalence of severe infections (§4.6), the attack rate for severe infections (equation (4.19)) and the equilibrium prevalence of severe infections (equation (4.5c)). These results are illustrated for parameters that are representative of four variants of SARS-CoV-2 in figures 4–6.
— Figure 4 shows that if masking primarily influences the probability that an infection is mild (\( m \)), then more effective masking strongly affects transmission (reducing \( R_0 \)) and lengthens the doubling time substantially, especially for less transmissible variants. In addition, the peak prevalence of severe cases is strongly affected (and the effect on peak prevalence is substantially greater for more transmissible variants). The expected number of severe cases during the initial wave of infections is also strongly dependent on \( m \), but is not sensitive to transmissibility over the range of \( R_0 \) observed for SARS-CoV-2 variants. The equilibrium prevalence of severe cases also declines with \( m \) (equations (4.1) and (4.5c)).

— Figures 5 and 6 show the effects of greater transmissibility of mild infections (increasing \( \beta_{\text{m}} \)), and longer infectious periods of mild cases (increasing \( 1/\gamma_{\text{m}} \)), respectively. The effects are similar, as might be expected given that \( R_0 \sim \beta/\gamma \) (equation (4.1)).

Overall, increasing the effectiveness of mask-induced variolation—whether by increasing the probability that an infection will be mild, reducing the transmissibility of mild infections or reducing the length of mild infections—has the potential to drastically impact disease control, by slowing spread and reducing the magnitude of the epidemic peak (‘flattening the curve’) [59,60], reducing the number of severe cases in the initial wave and reducing the prevalence of severe cases at equilibrium.

6.2. Limitations

The principal limitation of our analysis is that the hypothetical variolating effect of facemasks is unproven. We do not know that masking does tend to induce milder infections nor, if so, that variolated individuals attain a similar level of immunity as those who are infected in the absence of masks. Comments [61] and responses [62–64] to Gandhi & Rutherford’s initial article [17] make clear that further experimental and observational research is required. It is also worth noting that even if mask-induced variolation were very effective, promoting it could ‘implicitly encourage reckless behaviour’ [63] and consequently could present additional challenges for public health messaging.

6.3. Conclusion

Beyond qualitative conclusions, we are not likely to be able to make more powerful inferences without experimental studies that convincingly quantify the magnitudes of the effects that induce variolation from masking. If such experimental data do become available—and support the hypothesis that masking induces a substantial variolating effect—it will then be worth expanding our simple model (3.1) to include explicit latent periods, hospitalization, age and social structure (e.g. schools, workplaces) and heterogeneities in adherence to masking and other control measures. With appropriate data and more realistic models, we may be able to make quantitative inferences that could usefully inform policy decisions.

In the context of the highly transmissible Delta and Omicron variants [39–42], and the potential evolution of new SARS-CoV-2 variants that are even more transmissible and/or more successfully evade existing vaccines [65], a better understanding of the effectiveness of masking in promoting variolation could be of great value. At the time of writing, vaccines for children under 5 are not yet approved [66,67], but approval is expected soon [68,69]. While vaccine availability for people of all ages is imperative, substantial vaccine hesitancy [70] and breakthrough infections among the vaccinated [71–74], make achieving herd immunity through vaccination an unattainable target at present. If that situation persists, potential mask-induced variolation could contribute to COVID-19 control as we transition to endemicity.

Data accessibility. This article has no additional data.

Authors’ contributions. Z.L.: formal analysis, funding acquisition, investigation, resources, software, visualization, writing—original draft, writing—review and editing; D.E.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. Both authors received funding from the Natural Sciences and Engineering Research Council of Canada (NSERC): Z.L. was supported by an Undergraduate Student Research Award, and DJDE was supported by a Discovery Grant and an Emerging Infectious Diseases Modelling Initiative grant.

Endnote

The equivalence of the expressions for the discriminant in (4.8) and (4.9) can be verified directly, or by recognizing that the results must be invariant to changing the labelling of mild and severe: \( m \rightarrow s, \ s \rightarrow m, \ m \rightarrow (1 - m) \).

References

1. Dr Vivek Murthy, U S Surgeon General. 2020 The best way to protect yourself and your community is with everyday preventive actions, like staying home when you are sick and washing hands with soap and water, to help slow the spread of respiratory illness. Get your FluShot—fewer flu patients = more resources for COVID19. (accessed 15 July 2021). See https://twitter.com/Surgeon_General/status/1233726563881029632.

2. World Health Organization. 2020 Advice on the use of masks in the context of COVID-19: interim guidance, 6 April 2020. (accessed 15 July 2021). See https://apps.who.int/iris/handle/10665/331693.

3. Khazan O. 2020 Why we are running out of masks. See https://www.theatlantic.com/health/archive/2020/04/why-were-running-out-of-masks-in-the-coronavirus-crisis/609757/.

4. Sun L. 2020 Face mask shortage prompts CDC to loosen coronavirus guidance. Washington Post. See https://www.washingtonpost.com/health/2020/03/10/face-mask-shortage-prompts-cdc-loosen-coronavirus-guidance/.

5. Dyer E. 2020 The great PPE panic: how the pandemic caught Canada with its stockpiles down. See https://www.cbc.ca/news/politics/ppe-pandemic-covid-coronavirus-masks-1.5645120.

6. NPR. 2020 CDC Now Recommends Americans Consider Wearing Cloth Face Coverings In Public. (accessed 15 July 2021). See https://www.npr.org/sections/coronavirus-live-updates/2020/04/03/826219824/president-trump-says-cdc-now-recommends-americans-wear-cloth-masks-in-public.

7. Britneff B. 2020 Coronavirus: wearing non-medical masks will help others but not you, Tam says—National — Globalnews.ca. Global News. See https://www.globalnews.ca/news/6782934/coronavirus-non-medical-masks-community-covid-19/.

8. So E, Baker H. 2020 Mask or no mask? A look at UK’s policy over time. Centre for Research in the
59. The Economist. 2020 Covid-19 is now in 50 countries and things will get worse. See https://www.economist.com/briefing/2020/02/29/covid-19-is-now-in-50-countries-and-things-will-get-worse.

60. Centers for Disease Control and Prevention. 2007 Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States: early, targeted, layered use of nonpharmaceutical interventions. See https://stacks.cdc.gov/view/cdc/11425.

61. Klompas M, Morris CA, Sinclair J, Pearson M, Shenoy ES. 2020 Universal masking in hospitals in the Covid-19 Era. N. Engl. J. Med. 382, e63. (doi:10.1056/NEJMp2006372)

62. Rasmussen A, Escandón K, Popescu S. 2020 Facial Masking for Covid-19. N. Engl. J. Med. 383, 2092. (doi:10.1056/NEJM20200886)

63. Brosseau L. 2020 Facial Masking for Covid-19. N. Engl. J. Med. See https://www.nejm.org/doi/full/10.1056/NEJM20200886.

64. Gandhi M. 2020 Facial Masking for Covid-19. N. Engl. J. Med. See https://www.nejm.org/doi/full/10.1056/NEJM20200886.

65. Otto SP et al. 2021 The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. Curr. Biol. 31, R918–R929. (doi:10.1016/j.cub.2021.06.049)

66. Centers for Disease Control and Prevention. 2022 COVID-19 Vaccines for Children and Teens. See https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations/children-teens.html.

67. Public Health Agency of Canada. Vaccines for the COVID-19 Era. See https://www.canada.ca/en/public-health/services/vaccination-children/covid-19.html.

68. Johnson C. 2022 What to know about the coronavirus vaccine for children under 5. See https://www.washingtonpost.com/health/2022/01/11/covid-vaccine-kids-under-5/.

69. Bever L. 2022 Pfizer and BioNTech seek emergency use authorization of the first coronavirus vaccine for children under 5. See https://www.washingtonpost.com/health/2022/02/01/coronavirus-vaccine-under-5-pfizer/.

70. Bonyton S. 2021 Nearly 20% of Canadians still hesitant or refusing to get COVID-19 vaccine: poll. (accessed 7 September 2021). See https://globalnews.ca/news/7960345/covid-canada-vaccine-hesitancy-poll/.

71. Public Health Ontario. 2021 Confirmed Cases of COVID-19 Following Vaccination in Ontario: 14 December 2020 to 21 August 2021. (accessed 7 September 2021). See https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-epi-confirmed-cases-post-vaccination.pdf?la=en.

72. Zhou R et al. 2022 Vaccine-breakthrough infection by the SARS-CoV-2 omicron variant elicits broadly cross-reactive immune responses. Clin. Transl. Med. 12, 720.

73. Chen J, Wang R, Gilby NB, Wei GW. 2022 Omicron Variant (B.1.1.529): infectivity, vaccine breakthrough, and antibody resistance. J. Chem. Inf. Model. 62, 412–422. (doi:10.1021/acs.jcim.1c01451)

74. Dimeglio C, Migueres M, Mansuy JM, Saivin S, Miedougé M, Chapuy-Regaud S. 2022 Antibody titers and breakthrough infections with Omicron SARS-CoV-2. J. Infect. 84, e13–e15. (doi:10.1016/j.jinf.2022.01.044)