

Estimating the impact of COVID-19 control measures using a Bayesian model of physical distancing

Sean C. Anderson¹, Andrew M. Edwards^{1,2}, Madi Yerlanov³, Nicola Mulberry³,
Jessica E. Stockdale³, Sarafa A. Iyaniwura⁴, Rebeca C. Falcao⁵, Michael C.
Otterstatter^{5,6}, Michael A. Irvine⁷, Naveed Z. Janjua^{5,6}, Daniel Coombs⁴, and
Caroline Colijn^{*3}

¹Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, Canada

²Department of Biology, University of Victoria, Victoria, BC, Canada

³Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada

⁴Department of Mathematics and Institute of Applied Mathematics, University
of British Columbia, Vancouver, BC, Canada

⁵British Columbia Centre for Disease Control, Vancouver, BC, Canada

⁶School of Population and Public Health, University of British Columbia,
Vancouver, BC, Canada

⁷British Columbia Children's Hospital Research Institute, Vancouver, BC,
Canada

April 17, 2020

*Corresponding author: ccolijn@sfu.ca

Abstract

Extensive physical distancing measures are currently the primary intervention against coronavirus disease 2019 (COVID-19) worldwide. It is therefore urgent to estimate the impact such measures are having. We introduce a Bayesian epidemiological model in which a proportion of individuals are willing and able to participate in distancing measures, with the timing of these measures informed by survey data on attitudes to distancing and COVID-19. We fit our model to reported COVID-19 cases in British Columbia, Canada, using an observation model that accounts for both underestimation and the delay between symptom onset and reporting. We estimate the impact that physical distancing (also known as social distancing) has had on the contact rate and examine the projected impact of relaxing distancing measures. We find that distancing has had a strong impact, consistent with declines in reported cases and in hospitalization and intensive care unit numbers. We estimate that approximately 0.78 (0.66–0.89 90% CI) of contacts have been removed for individuals in British Columbia practising physical distancing and that this fraction is above the threshold of 0.45 at which prevalence is expected to grow. However, relaxing distancing measures beyond this threshold re-starts rapid exponential growth. Because the extent of underestimation is unknown, the data are consistent with a wide range in the prevalence of COVID-19 in the population; changes to testing criteria over time introduce additional uncertainty. Our projections indicate that intermittent distancing measures—if sufficiently strong and robustly followed—could control COVID-19 transmission, but that if distancing measures are relaxed too much, the epidemic curve would grow to high prevalence.

1 Introduction

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has now spread worldwide and resulted in over 1.5 million diagnosed cases and 100,000 deaths globally [1]. Estimates of COVID-19 case fatality rates from Hubei, China, the rest of China, and other countries range from 0.3% to above 5% in different populations at various times [2, 3], with an estimate of 1.4% [4] currently favoured in some analyses [5]. Age [6] and comorbidities (hypertension, cardiovascular and respiratory disease [7]) are strong risk factors for severe illness, hospitalization, and death. Furthermore, COVID-19 poses severe challenges for health care, with risks that requirements will exceed hospital bed, critical care and ICU capacities even in well-resourced health care systems. In the current absence of a vaccine or effective therapeutic options, widespread physical distancing measures (also known as social distancing measures) and handwashing are the main interventions currently available to reduce transmission. Countries have used a variety of such physical or social distancing measures including cancelling mass gatherings, closing restaurants, work-from-home orders, and “lockdowns” of varying strictness.

In British Columbia (BC), Canada, the first case of infection was detected on January 26, 2020 with sporadic cases related to travel until March 8, followed by a sustained increase in cases. A number of measures were implemented over the following weeks to reduce transmission (Figure 1). However, the direct impact of these measures on transmission is not known. Distancing measures have high economic, health, and social impacts. Thus, there is an urgent need to understand what level of contact rate and physical distancing measures are optimal to reduce transmission. Once initial transmission has been brought under control, as in China and Korea [8, 9],

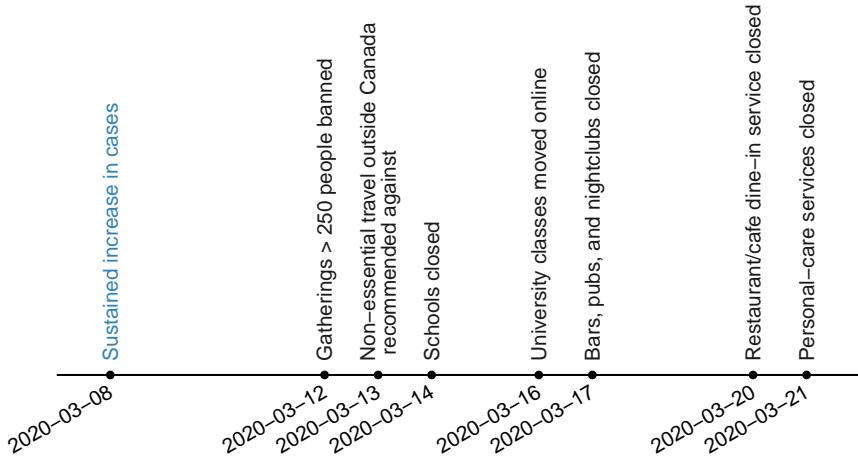


Figure 1: Key physical distancing measures implemented in British Columbia, Canada in response to COVID-19. Schools closed for an annual two-week break on March 14, then were declared indefinitely closed on March 17.

there remains the question of what relaxation in social measures could keep transmission under control.

There have been a number of models simulating the impact of broad physical distancing measures [8, 10, 11, 12]. Direct estimates of the strength and impact of distancing measures have focused on the effective reproduction number over time, using approaches based on reported deaths [13] or confirmed cases [14]. These estimates are influenced by the assumed serial interval distribution and the infection fatality and delay between symptom onset and death. In comparisons among EU countries, estimates assume that physical-distancing measures impact each location equally and that all deaths are reported [13]. Inferences from reported deaths also impacted by the assumption that deaths are independent, whereas in British Columbia, more than half are associated with clusters in long term care facilities. Estimates of the effective reproduction number based on reported cases have been adjusted for the delay between symptom onset and reporting [14, 15], but cannot accommodate underestimation or asymptomatic or weakly symptomatic individuals. Furthermore, the effective reproduction number is a broad summary of the overall growth of the epidemic, and is not a direct estimate of the impact of physical distancing on the contact patterns relevant to transmission.

Here, we introduce an epidemiological model of physical distancing and assess the degree to which contact rates have changed—for the population that is participating in physical distancing—due to recent policy measures in British Columbia. The model structure is informed by a survey of attitudes and responses to messaging about COVID-19 and physical distancing. We use an observation model to reflect the relationship between case counts and incident infections and to account for underestimation of the time between symptom onset and case reporting. We then use a Bayesian approach to estimate the reduction in contact among those participating in physical distancing measures, using reported case counts for the period March 1 to April 11, 2020. We explore to what extent distancing measures could be relaxed while keeping case numbers under control.

2 Methods

2.1 Data

We fit the physical distancing model to case count data from British Columbia from March 1, 2020 (when a total of eight cases had been detected in the province) to April 11, 2020 at which time 1445 cases had been confirmed. These data are available in press releases from the BC Centre for Disease Control (BCCDC) [16], from the public data dashboard [17], and from the [code repository](#) associated with this paper. Testing procedures were adapted over the course of the outbreak. In particular, lab testing criteria were changed on March 16 to focus on hospitalized patients, healthcare workers, long term care facility residents/staff, and those people part of an existing cluster and experiencing respiratory symptoms. This led to instability in case counts in the surrounding days with some large jumps in the number of identified cases, which we correct for in the model. There was also considerable variability in the daily testing rate. During March, the daily number of completed tests ranged from approximately 100 to 3500, and did not strictly increase over time.

For some of the confirmed cases, estimates of the date of symptom onset are available (Figure 2). We used the delays between symptom onset and cases being reported to parameterize the physical distancing model. In this case-specific data set there are only seven cases reported before February 29, and a decline in reported cases after April 2 (Figure 2). Therefore we used only the 535 cases in the case-specific data that were reported between these dates to parameterize the delay part of the model.

The Angus Reid Institute conducted a survey to examine how physical distancing measures have changed behaviour (March 20–23, 2020; n = 1664; [18]). The survey questions included whether respondents believed the COVID-19 epidemic was “overblown”. Responses about behavioural practices around hand washing, keeping extra distance, and other distancing behaviours were stratified according to this belief.

2.2 Epidemiological model

We developed an epidemiological model to describe the numbers of individuals who are: susceptible (S); exposed to the virus (E_1); exposed, pre-symptomatic, and infectious (E_2); symptomatic and infectious (I); quarantined (Q); and recovered or deceased (R) over time (Figure 3). Recovered individuals are assumed to be immune. The model includes analogous variables for individuals practising physical distancing, represented by subscript d, i.e. S_d , E_{1d} , E_{2d} , I_d , Q_d , and R_d . Physical distancing reduces the frequency of contact between individuals (and hence between compartments; Figure 3).

Figure 2 consists of three panels (A, B, C) illustrating COVID-19 data from March 1 to April 11, 2020.

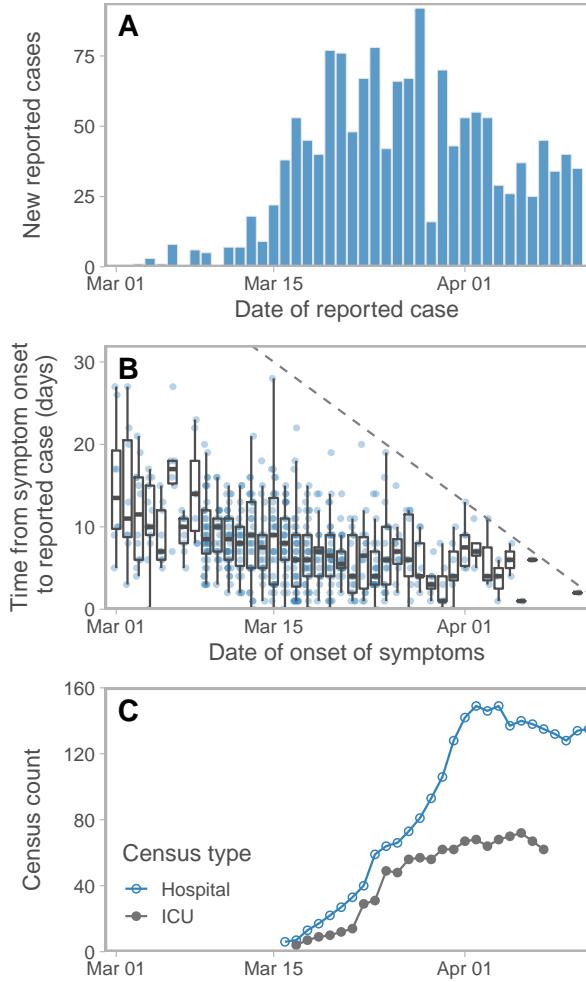


Figure 2: (A) New reported cases per day and (B) time from symptom onset to reporting from the case-specific data as of April 11, 2020. The dashed line in panel B represents the 1:1 line. Cases above this line, by definition, have not been reported yet. (C) Hospitalization and ICU (Intensive Care Unit) census counts. All data are from the BC Centre for Disease Control.

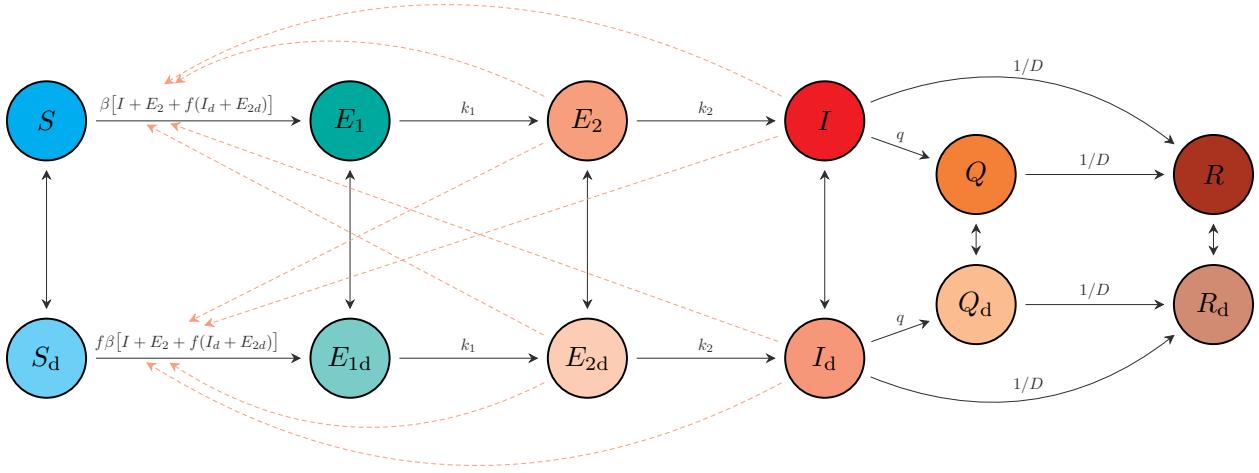


Figure 3: Schematic of model described by Equations (1) and (2). An individual in some compartment X can begin distancing and move to the corresponding compartment X_d at rate u_d . The reverse transition occurs at rate u_r . The model quickly settles on a fraction $e = u_d/(u_d + u_r)$ participating in distancing, and dynamics depend on this fraction, rather than on the rates u_d and u_r .

The non-physical-distancing differential equations are:

$$\begin{aligned}
 \frac{dS}{dt} &= -\beta [I + E_2 + f(I_d + E_{2d})] \frac{S}{N} - u_d S + u_r S_d \\
 \frac{dE_1}{dt} &= \beta [I + E_2 + f(I_d + E_{2d})] \frac{S}{N} - k_1 E_1 - u_d E_1 + u_r E_{1d} \\
 \frac{dE_2}{dt} &= k_1 E_1 - k_2 E_2 - u_d E_2 + u_r E_{2d} \\
 \frac{dI}{dt} &= k_2 E_2 - qI - \frac{I}{D} - u_d I + u_r I_d \\
 \frac{dQ}{dt} &= qI - \frac{Q}{D} - u_d Q + u_r Q_d \\
 \frac{dR}{dt} &= \frac{I}{D} + \frac{Q}{D} - u_d R + u_r R_d,
 \end{aligned} \tag{1}$$

where β is the transmission parameter, D is the mean duration of the infectious period, f represents physical distancing, u_d and u_r are the rates of movement to and from the physical distancing compartments, k_1 is the rate of movement from the E_1 to E_2 compartment, k_2 is the rate of movement from the E_2 to I compartment, and q is the quarantine rate for movement from the I to Q compartment. The physical distancing compartments contribute a reduced amount (a fraction f) to the force of infection, with $f = 1$ representing no physical distancing. We make f time-dependent (see below) to represent changes in physical distancing. We report the estimated impact of distancing in terms of the fraction of contacts that are reduced (i.e., $1 - f$), so that “strong distancing” is a high fraction (or percent), and “weak distancing” is a low fraction (or percent). In the model without interventions (neither distancing nor quarantine), the basic

reproductive number R_{0b} is $\beta(D + 1/k_2)$, namely the transmission rate times the mean duration in the infectious compartment.

There are six analogous equations for the physical distancing compartments (denoted by subscript d). The physically distancing compartment S_d has a contact rate with infectious individuals that is a fraction f of that for the S compartment. This factor appears twice: distancing individuals contribute less to the force of infection and also are less likely to encounter others than non-distancing individuals, giving

$$\begin{aligned} \frac{dS_d}{dt} &= -f\beta [I + E_2 + f(I_d + E_{2d})] \frac{S_d}{N} + u_d S - u_r S_d \\ \frac{dE_{1d}}{dt} &= f\beta [I + E_2 + f(I_d + E_{2d})] \frac{S_d}{N} - k_1 E_{1d} + u_d E_1 - u_r E_{1d} \\ \frac{dE_{2d}}{dt} &= k_1 E_{1d} - k_2 E_{2d} + u_d E_2 - u_r E_{2d} \\ \frac{dI_d}{dt} &= k_2 E_{2d} - q I_d - \frac{I_d}{D} + u_d I - u_r I_d \\ \frac{dQ_d}{dt} &= q I_d - \frac{Q_d}{D} + u_d Q - u_r Q_d \\ \frac{dR_d}{dt} &= \frac{I_d}{D} + \frac{Q_d}{D} + u_d R - u_r R_d. \end{aligned} \tag{2}$$

We model the change in physical distancing that occurred in British Columbia in March 2020 by a simple linear function whereby physical distancing increases ($f(t)$ decreases) over one week between 15th March ($t = t_1$) and 22nd March ($t = t_2$):

$$f(t) = \begin{cases} 1, & t < t_1, \\ f_2 + \frac{t_2 - t}{t_2 - t_1}(1 - f_2), & t_1 \leq t < t_2, \\ f_2, & t \geq t_2. \end{cases} \tag{3}$$

We fix $f(t) = f_2$ until the final day of observed data and change $f(t)$ for future days to represent different scenarios regarding relaxation of physical distancing.

The number of reported cases each day is C_r , where r represents discrete days (whereas t is continuous time in the model). We also have data on each individual case, some of which have an estimated date of the onset of symptoms. The number of people per day who become symptomatic is the number per day moving from the E_2 and E_{2d} compartments to the I and I_d compartments.

Due to the delay between symptom onset and reporting, the model's predicted number of reported cases, μ_r , on integer day r is comprised of contributions from previous days weighted by a delay:

$$\mu_r = \psi_r \int_0^{45} k_2 [E_2(r-s) + E_{2d}(r-s)] w(s) ds, \tag{4}$$

where $w(s)$ is the density function for delay s and 45 is the maximum delay that needs to be considered (see Supplement). Also, ψ_r is the proportion of anticipated cases on day r that are tested and reported; this varies over time due to changes in the testing protocols, lab capacity and other factors. In our data, the number of tests performed each day jumped dramatically on

Table 1: Parameters and sources. The duration of the infectious period is shorter than the duration of severe illness, accounting for self-isolation and less severe illnesses. The quarantine parameter q reflects approximately 1/5 of severe cases either ceasing to transmit due to hospitalization or completely self-isolating. The model depends on the combination $u_r/(u_r + u_d)$, the fraction engaged in physical distancing, estimated from the survey data cited above. The testing patterns have changed over time, with laboratories increasing the numbers of tests on approximately March 14 (motivating our change in ψ_r).

Symbol	Definition	Default value	Justification
N	Population of British Columbia	5,100,000	[21]
D	Mean duration of the infectious period	5 days	[22, 23]
k_1	(time to infectiousness) $^{-1}$ (E_1 to E_2)	0.2 days $^{-1}$	[24, 25, 26]
k_2	(time period of pre-symptomatic transmissibility) $^{-1}$ (E_2 to I)	1 days $^{-1}$	[25, 26]
q	Quarantine rate	0.05	[27]
u_d	Rate of people moving to physical distancing	0.1	[18]
u_r	Rate of people returning from physical distancing	0.02	[18]
ψ_r	Proportion of anticipated cases on day r that are tested and reported	0.1 0.3	Before March 14 On and after March 14

March 14; we model this with a sharp increase in ψ_r on that date (0.1 to 0.3) and attempt to fit a random walk to this function as a supplementary analysis.

We use a Weibull distribution for the delay function $w(s)$, and fit the shape and scale parameters using the case-specific data of reported cases and time of symptom onset (Figure 2B). In the Supplement we derive a novel multinomial likelihood function (based on [19] and [20]) to handle right-truncation. The resulting maximum likelihood estimate (with 95% univariate confidence interval) of the shape is 1.73 (1.60–1.86) and of the scale is 9.85 (9.30–10.46). Using the maximum likelihood estimates, the mean delay from symptom onset to reporting is 8.78 days, one day longer than the data's mean (7.79 days) due to accounting for right-truncation (Figure S2).

2.3 Fixed parameter and variable values

The nature of the epidemiological model requires us to assume a number of parameter and variable values (Table 1). We test the robustness of our results to the key fixed values as supplementary analyses. Specifying initial conditions for the epidemiological model involves setting a total number of infected people I_0 at an initial point in time (February 1, 2020), which are divided among the exposed and infected compartments. We define $e = u_d/(u_d + u_r)$ to be the fraction practising physical distancing ($e = 0.83$ in the base model). To further avoid sharp initial transient behaviour we also initially distribute all individuals among the distancing and non-distancing components, to give reasonable conditions for March 1 (Table 2). The model begins on February 1 with 8 cases, reflecting a 10–30% chance of detection [5] and very low reported cases at that time. We do not model introductions during February, and instead we compensate for this with an elevated initial number of infectious cases.

Table 2: Initial values of variables for February 1, 2020.

Non-distancing			Distancing		
Variable	Initial definition	Initial value	Variable	Initial definition	Initial value
S	$(1 - e)(N - I_0)$	849999	S_d	$e(N - I_0)$	4249993
E_1	$0.4(1 - e)I_0$	0.53	E_{1d}	$0.4eI_0$	2.67
E_2	$0.1(1 - e)I_0$	0.13	E_{2d}	$0.1eI_0$	0.67
I	$0.5(1 - e)I_0$	0.67	I_d	$0.5eI_0$	3.33
Q	0	0.00	Q_d	0	0
R	0	0.00	R_d	0	0

2.4 Observation model

We used a Bayesian statistical model to condition our inference about R_{0b} , f_2 , and expected case counts μ_r on the observed case count data C_r . We related the expected number of cases to the observations through a negative binomial observation model using the NB2 parameterization of the probability mass function [28]:

$$\text{NB2}(C_r | \mu_r, \phi) = \binom{C_r + \phi - 1}{C_r} \left(\frac{\mu_r}{\mu_r + \phi}\right)^{C_r} \left(\frac{\phi}{\mu_r + \phi}\right)^\phi, \text{ for } r = 1, \dots, 42, \quad (5)$$

in which the variance scales quadratically with the mean according to the (inverse) dispersion parameter, ϕ : $\text{Var}[C] = \mu + \mu^2/\phi$. Using the notation $[a|b]$ to describe the conditional distribution of a given b and the bold symbol \mathbf{C} to represent the vector of cases, we can write the joint posterior distribution for the model as

$$[R_{0b}, f_2, \phi | \mathbf{C}] \propto [\mathbf{C} | R_{0b}, f_2, \phi][R_{0b}][f_2][\phi], \quad (6)$$

where the term $[\mathbf{C} | R_{0b}, f_2, \phi]$ refers to the negative binomial data likelihood and the terms $[R_{0b}]$, $[f_2]$, and $[\phi]$ refer to the prior distributions.

We placed weakly informative priors on R_{0b} , f_2 , and ϕ . We chose a prior on R_{0b} to encompass values commonly published in the literature for SARS-CoV-2 [2, 29]: Lognormal(log(2.6), 0.2). We chose a prior for f_2 that resulted in a mean of 0.4 and a standard deviation of 0.2, $f_2 \sim \text{Beta}(2, 3)$, to represent a moderately strong reduction in contact fraction while still being broad enough to encompass a wide range of values. We chose a prior on ϕ to constrain the model to avoid substantial prior mass on a large amount of over-dispersion (small values of ϕ): $1/\sqrt{\phi} \sim \text{Normal}(0, 1)$ [30].

We fit our models with Stan 2.19.1 [31, 32] and R 3.6.2 [33]; Stan implements the No-U-Turn Hamiltonian Markov chain Monte Carlo algorithm [34] for Bayesian statistical inference. In our main model run, we sampled from eight chains with 2000 iterations each and discarded the first 1000 iterations of each chain as warm-up. We assessed chain convergence visually via trace plots (Figure S7) and via ensuring that $\hat{R} < 1.01$ (the potential scale reduction factor) and that ESS > 200 (the effective sample size), as calculated by the rstan R package [32]. We represented uncertainty via quantile-based credible intervals.

2.5 Model validation

To test the model's ability to recover the parameters and to test our implementation of the Stan model, we repeatedly fit the model to data simulated from separate code in R. We inspected the resulting posterior distributions for bias and coverage of the known true values and visually inspected the resulting time series (e.g., Figures S3 and S4). Furthermore, we conducted posterior predictive checks [35] to assess whether the observed data were consistent with data generated by the model. Code to reproduce our analysis is available at <https://github.com/carolinecolijn/distancing-impact-covid19>.

3 Results

We find that physical distancing has considerably reduced the contact rate. Our estimate of the fraction of contacts that have been removed due to distancing ($1 - f_2$) is 0.78 (0.66–0.89 90% CI [credible interval]), which is somewhat above the critical threshold (0.45; Figure 4). Figure 4 shows the case counts with the model fit, along with the posterior estimates for R_{0b} and the fraction of contacts that have been removed due to distancing measures. The model describes the count data well, with reported cases showing a peak in late March, approximately two weeks after the initiation of distancing measures (Figure 4A). The data are informative with respect to both main parameters, with the posteriors distinctly different and more peaked than the priors (Figure 4B,D). Figures S5–S7 show posterior samples for the model variables, examples of predicted case counts, and trace plots illustrating MCMC convergence. Figure S8 shows the joint posterior scatterplot. We have used a fixed value of e , the fraction engaged in distancing; this choice is motivated by the survey and behavioural data. If e were lower, the estimated strength of distancing would be higher to achieve the same case counts. We find this tradeoff analytically using the basic reproduction number for the full model (Supplement, Figure S1).

We find that with a shorter incubation period and duration of infectiousness, a lower reproduction number will fit the same overall growth rate, and conversely if the infectious duration (and serial interval) are higher, a higher reproduction number is required but the fit to data is very similar (Figure S10). This relationship is well known [36]. The conclusion that distancing measures have reduced contact is robust to these alternatives (Figure S10). The model depends on the fraction engaged in distancing, but not much on the rates u_d and u_r ; Figure S11 illustrates that we obtain the same results with these rates increased by a factor of 10. We also explore the robustness to the unknown underestimation fractions (Figure S12), and a random walk pattern in the fraction of cases sampled (Figure S13); again, the data are consistent with a range of underestimation fractions, and the conclusion about the contact reduction is robust.

Our estimates suggest that some relaxation of current distancing measures would be possible without bringing the growth rate above zero (Figure 5A,D), but if measures are relaxed too much (in the absence of monitoring and re-starting measures), the prevalence and case counts begin to increase exponentially (Figure 5B–C,E–F), reaching high levels by June 2020 if distancing were to cease entirely (Figure 5F; Figure 6). These are illustrative scenarios only; public health responses with renewed or revised measures would likely be put in place rapidly were such rises to be observed. The speed of growth depends on how close the system is to the epidemic threshold

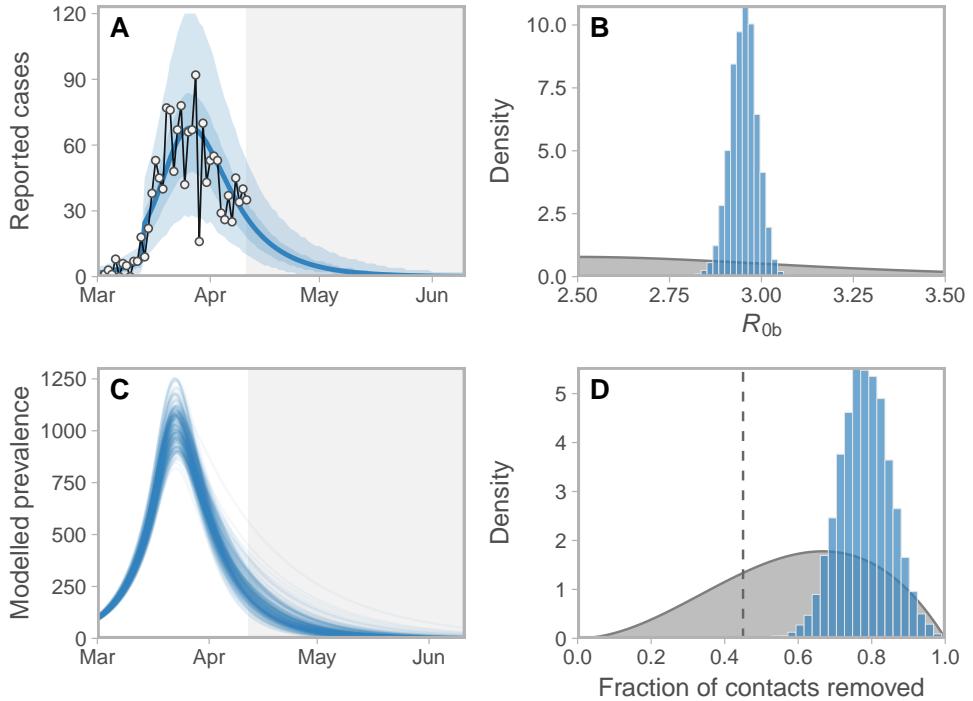


Figure 4: (A) Observed and estimated case counts, (C) estimated prevalence, and posterior estimates for (B) R_{0b} and (D) fraction of contacts removed ($1 - f_2$). These projections do not account for introduced cases from other jurisdictions and they assume that distancing measures remain in place. The fraction of contacts removed is the model's portion of contacts that have been stopped among those who are engaged in physical distancing, with 0.0 meaning no change and 1.0 meaning 100% of contacts have been removed for those engaged in physical distancing. In panel A, the blue line represents the posterior mean and the shaded ribbons represent 50% and 90% credible intervals on new observations. Dots and black lines represent the reported data. In panel C, lines represent example draws from the posterior. In panels B and D, priors are shown in grey and posteriors in blue. In panel D, the dashed vertical line denotes the threshold below which an exponential increase in prevalence is expected (see Figure S9). Note: Model prevalence depends on our assumptions about underestimation, incubation period, and the duration of infection, none of which we can estimate well from these data (Figures S12). This model does not take asymptomatic individuals into account. Much higher values of the prevalence are consistent with our data.

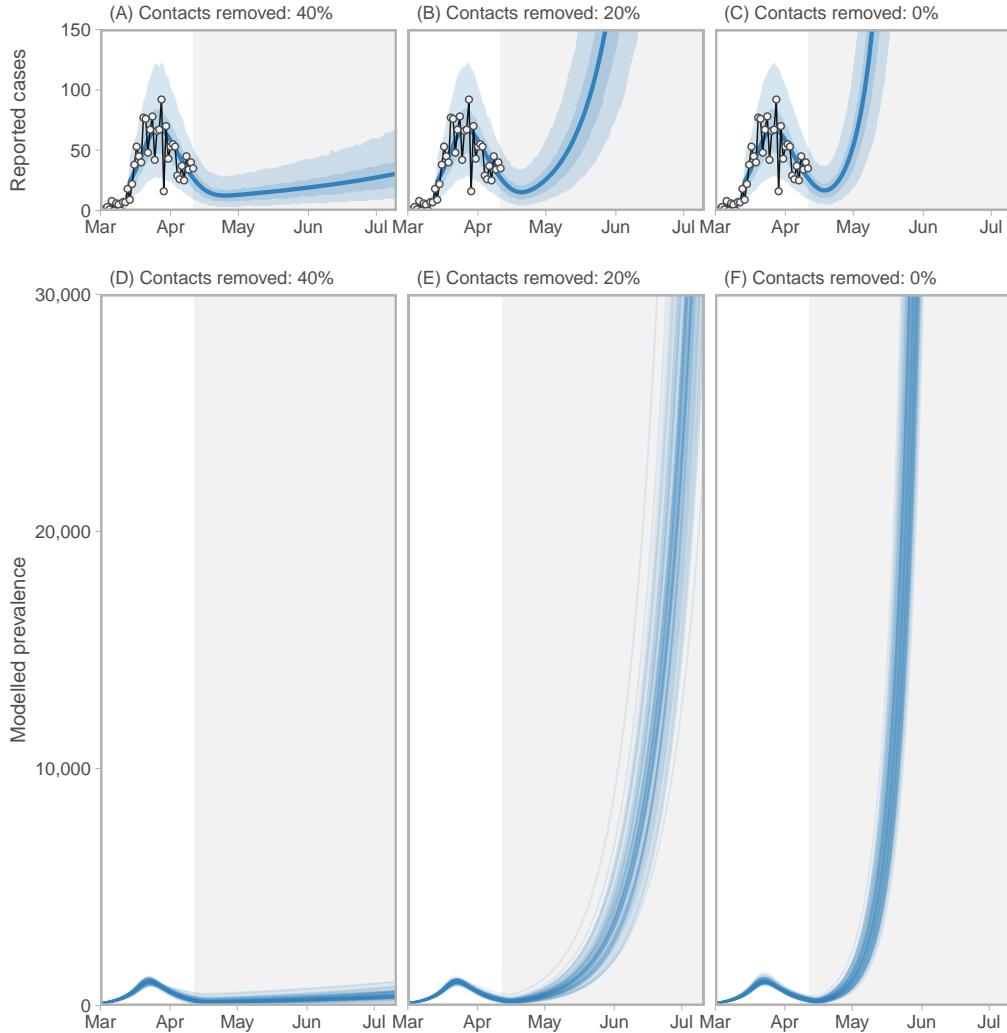


Figure 5: Three scenarios for relaxing distancing measures beginning on April 12 2020. (A, D) Distancing measures are relaxed but prevalence and reported cases continue to decline, with slower decline and more variance in the number of reported cases in coming weeks. (B-C, E-F) Distancing measures are relaxed (B, E) or cease (C, F) and exponential growth is observed beginning mid-April. Figure description is otherwise the same as for Figure 4A, C. Note: Model prevalence depends on our assumptions about underestimation, incubation period, and the duration of infection, none of which we can estimate well from these data (Figures S12). This model does not include individuals who do not develop symptoms. Much higher values of the prevalence are consistent with our data.

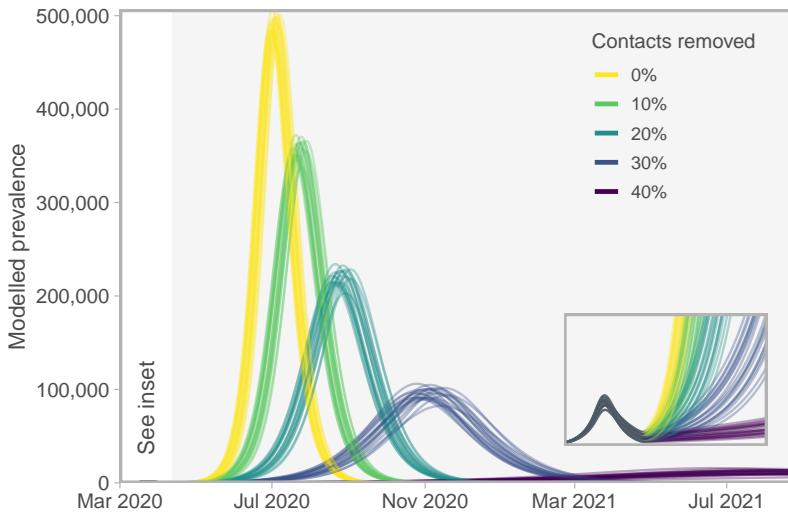


Figure 6: Epidemiological curves for five scenarios of relaxing distancing measures (beginning on April 12 2020) that are insufficient to keep the prevalence from growing. Stronger measures “flatten the curve”, as would be expected. Lines represent 20 draws from the joint posterior distribution. Grey region denotes future dates during which the relaxed distancing measures are applied. Inset box magnifies the lower-left region of the plot to show the modelled prevalence from March 1 to July 1 2020—the initial curve is imperceptible on the main axes. **Note:** Model prevalence depends on our assumptions about underestimation, incubation period, and the duration of infection, none of which we can estimate well from these data (Figures S12). This model does not include individuals who do not develop symptoms. Much higher values of the prevalence are consistent with our data.

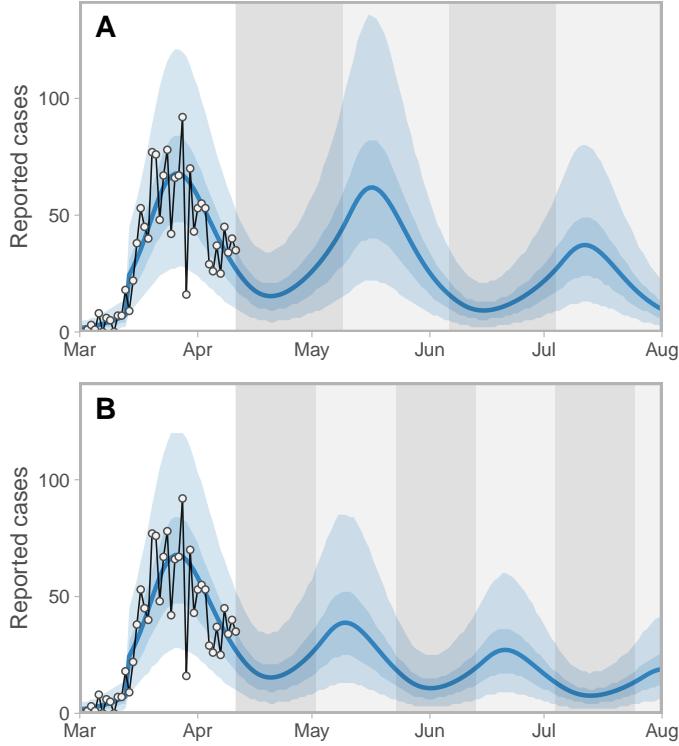


Figure 7: Two scenarios of cycling between physical distancing levels. Here, the percentage of contact reduction alternates between 20% and 78% at (A) four-week and (B) three-week intervals. Reducing contacts by 78% is approximately the estimated level by our model (Figure 4). Dark- and light-grey shaded regions represent the 20% and 78% time periods, respectively. Note the lag between changes in physical distancing and reported case counts. Blue lines represent the posterior mean and shaded ribbons represent 50% and 90% credible intervals on new observations. Dots and black lines represent the reported data.

(Figure S9). If strong enough measures are not maintained, the model predicts a range of possible epidemic curves (Figure 6) consistent with published simple and complex models [11, 10, 13].

There has been some interest in relaxing distancing measures and re-introducing them when a threshold has been reached, such as when intensive care capacity is nearly reached [37]. We did not explore a dynamic trigger, but did explore the behaviour when distancing measures are introduced and relaxed repeatedly (Figure 7). If the relaxation period is such that the outbreak remains contained throughout (with an effective reproduction number less than one), then the prevalence declines at alternating faster and slower rates. Simulations in which the relaxation of distancing measures allows prevalence to rise again are of more interest. In a scenario of switching between the current mean estimate (78% of contacts removed) and 20% contacts removed, reported cases rise, lagging the relaxation of distancing (Figure 7). Illustrative simulations in which distancing alternates every three or four weeks allows an overall continued decline; however, the longer the period of relaxation, the more the prevalence is able to rise in between periods of distancing (Figure 7). Control of delayed feedback systems is challenging [38], and ideally if a dynamic trigger such as reported or ICU cases were to be used, monitoring would need to be as rapid as possi-

ble. Monitoring of distancing behaviour and population contact patterns would be important, in addition to monitoring cases.

4 Discussion

Our results suggest that physical distancing measures are effective in British Columbia at present. We estimate the fraction of contact reduction to be 0.78 (0.66–0.89 90% CI), which is stronger than the critical threshold of the model. Our conclusions about the decline in cases are supported by declines in hospitalizations and ICU admissions. Our estimate is in line with survey data from the UK [39] which suggested a 73% reduction in contacts, and a modelling study which found that a 70–80% reduction in contacts is consistent with data in France to date [12].

Our estimate of the effect of distancing on contact patterns is consistent with independent local lines of evidence for the strength of distancing measures. Local rail (Skytrain) station crowding data, provided by Metro Vancouver's transportation authority TransLink, gives a proxy for reduction in public transport use. Between the weeks of February 24 and March 23, travel from 7–10am was reduced by 73% (on weekdays) and 63% (on weekends), from 11am–4pm by 71% and 72%, from 5pm–8pm by 79% and 80%, and from 8pm–12am by 73% and 76%. Overall daytime travel was reduced by 16% for the week of March 9, 64% for the week of March 16, and 73% for the week of March 23. Estimates on adhering to physical distancing are also available from a publicly available respondent-driven survey [18]. The survey found a the rate of respondents stating that there was a serious threat of a coronavirus outbreak in Canada increased from 42% on March 5–6 to 88% on March 20–23. For individuals who stated there was a serious threat, 89% stated they were keeping personal distance compared to 66% for individuals who did not believe there was a serious threat. Mobile phone location data from BlueDot [40] suggested that the maximum and cumulative distance travelled from home fell by approximately 90%, the portion of mobile phone check-ins at home rose by over 10%, and the portion of devices for which every check-in was at home rose by 60%. These estimates are also consistent with Google mobility [41], Citymapper index [42], and Apple mobility data [43]. These are indirect reflections of the contact rate but are supportive of a dramatic change in contact patterns as reflected in our model.

Our results suggest that some relaxation of distancing measures may be possible; we have simulated fixed and dynamic measures, less stringent than the measures in place at present, that would continue to maintain low case numbers. This is feasible either through continual distancing that is quite strong, or via well-monitored dynamic on/off measures. We have illustrated the model's high case volumes and long time frames that would result from cessation of distancing and the absence of continued strong public health and behavioural intervention. These simulations illustrate that immunity has not built up in the model; our estimates of the decline are not due to a natural peak in an epidemic curve, but are the direct effect of changes in the contact patterns in the population. Seasonal transmission could even amplify a peak in the winter season, if control and monitoring were ceased then [11]. However, experience to date indicates that given any early signs of renewed growth, public health in BC is in a position to manage and reduce transmission, as evidenced by current declines in BC; the benefit of recent experience could enable the response to be swifter and more effective if case numbers begin to rise again.

Distancing measures were in place in BC from approximately March 15 and we believe that they had taken full effect by the week of 22 March. The peak in observed cases occurred at

approximately March 27, 12 days after the model's ramp-up of distancing began and 15 days after the initiation of any distancing measures. This delay presents obvious challenges for monitoring the success of control measures using reported case count data. We suggest that two different kinds of monitoring will be important if distancing measures are to be relaxed: (1) monitoring cases, through testing, contact tracing, and other case finding, and (2) monitoring contact patterns and distancing behaviour in the population, in a kind of "distancing surveillance" effort. These data, derived from mobile phones, surveys and apps, could be available very rapidly, whereas the incubation period places an unavoidable delay between control measures and detecting their impact in reported cases— even if testing were widespread and reporting were immediate. There is considerable interest in real-time monitoring of mobile phone movements, population surveys on the uptake of physical distancing, and other behavioural data. While these are potentially promising avenues, the outcome of interest is incident infections. Locations of mobile phones, traffic patterns, and survey information are all proxies for this outcome at best. The work we have presented here could help to calibrate distancing surveillance measurements, to understand how they relate to changes in contact rates for modelling efforts.

Our modelling framework has a number of important limitations. We do not model age and contact structure explicitly, except to distinguish between two populations: those participating in distancing or not. This has the advantage that we do not require data on age-specific contact patterns, responses to distancing measures, or infectiousness; these data are not available at this time. It also limits our ability to provide guidance on where and how contact reduction measures could be implemented. It is a simplification of behaviour in many ways; true distancing responses are a continuum, and the measures in place (e.g. no mass gatherings or dine-in services) also mean that the whole population is experiencing some changes in contact patterns. Our model is deterministic, and so does not capture the possibility of extinction; in addition, we have not simulated introductions of COVID-19 from other jurisdictions. We have not accounted for geographic structure; differences in distancing behaviour, health care practices, and demographics in different jurisdictions could impact the results. We have not modelled either conventional or automated contact tracing [44, 45]; in our model these would decrease the duration of the infectious period and change the transitions for some exposed individuals.

There are also limitations in our data. We have used an observation model to link reported cases to the modelled prevalence, and we included variation in the portion of cases detected over time. Modelling and forecasting based on reported cases faces challenges when testing is driven by clinical needs, testing capacities, and other constraints (and in particular is not designed to test population samples). Cases in long-term care facilities (LTCF) represent a substantial fraction of the cases, and particularly the deaths, in BC. Along with the low number of deaths in total, this is one rationale for not modelling deaths explicitly. We included LTCF cases but also modelled a wide range of under-reporting to account for potential biases. If many cases in an LTCF cluster were all reported on the same day (or within a short time frame) this could increase the noise in reported case counts. We have modelled case counts as over-dispersed to account for such variation.

The testing criteria include a number of categories that have changed over time. Testing volume increased sharply in mid March. In the latter half of March and until April 9, testing focused primarily on those hospitalized or likely to be hospitalized, health care workers, residents of long-term care facilities, and other cluster investigations. After April 9, testing was opened somewhat to include residents of remote, isolated or Indigenous communities, people who are homeless or

have unstable housing, and by physicians' clinical judgement. There is likely some inconsistency in the application of these guidelines across hospitals and facilities. As a consequence, the base population being tested changes with time and contributes a changing portion of the force of infection. Underestimation is therefore complex and is comprised of varying under-ascertainment and under-reporting.

There remain important unknowns about COVID-19 that give rise to additional limitations for modelling efforts; immunity and asymptomatic transmission are two of these. We have included pre-symptomatic transmission but we have not explicitly modelled asymptomatic individuals, who may have few or no symptoms but nonetheless be transmitting, and who may or may not be building lasting immunity. Recent work suggests that both pre- and asymptomatic individuals may be contributing considerably to transmission [46, 25, 26]. We have indirectly approached this uncertainty, exploring variable underestimation fractions and duration of the incubation and infectious periods. A wide range for underestimation and duration is consistent with the reported case data. Our conclusion about the impact of distancing measures appears to be robust to these uncertainties, although the basic reproductive number and the model prevalence vary according to which assumptions are made about underestimation and duration. Model predictions for the peak timing and size without strong public health interventions will depend strongly on the dynamics of immunity, including the numbers of asymptomatic individuals and their immunity [11].

Our model suggests that distancing measures are working in British Columbia, that some relaxation of these measures may be possible, and that this must be done carefully. Given the likely low levels of immunity, long-term public health measures will be necessary to control COVID-19. If data were available describing contact patterns and transmissibility by age, along with data describing the impact of specific activities on these contact patterns, then models could be effective tools to determine how to safely relax distancing measures—for example, restarting specific activities such as particular workplaces or schools. Without knowledge of prevalence and transmissibility in children, of the extent of asymptomatic infection, and of the contact patterns that would result from restarting specific activities, models aiming to simulate these activities will have large uncertainties. We therefore suggest that there is an urgent need for longitudinal measurements of population-level prevalence and immunity via viral testing and serological studies, even where prevalence is low, and despite this being challenging. This will inform our understanding of the role of children and schools. Furthermore, if distancing measures are to be relaxed, it will be crucial to have strong surveillance in place through widespread testing, contact tracing, and isolation of new cases, as well as strong compliance with potentially shifting public health policy and messaging.

Acknowledgements

We thank TransLink for providing data on transportation and Bluedot for descriptive summary statistics on mobile-phone mobility patterns. The research and analysis are based on data from TransLink and the opinions expressed do not represent the views of TransLink. The opinions and conclusions expressed also do not represent the views of Bluedot. We thank J.A. Otte for providing details on testing-criteria changes in BC and for insightful comments on an earlier version of this

manuscript. C.C. and J.S. are funded by the Federal Government of Canada's Canada 150 Research Chair program. We thank Fisheries and Oceans Canada for their support.

References

- [1] M. Roser, H. Ritchie, E. Ortiz-Ospina, [Coronavirus Disease \(COVID-19\) – Statistics and Research, Our World in Data](#).
URL <https://ourworldindata.org/coronavirus>
- [2] MIDAS Network, COVID-19, <https://github.com/midas-network/COVID-19>, [Online; last accessed April-2020].
- [3] D. D. Rajgor, M. H. Lee, S. Archuleta, N. Bagdasarian, S. C. Quek, The many estimates of the COVID-19 case fatality rate, *Lancet Infect. Dis.* In press. [doi:10.1016/S1473-3099\(20\)30244-9](https://doi.org/10.1016/S1473-3099(20)30244-9).
- [4] J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, G. M. Leung, Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China, *Nat. Med.* 26 (2020) 506–510. [doi:10.1038/s41591-020-0822-7](https://doi.org/10.1038/s41591-020-0822-7).
- [5] Using a delay-adjusted case fatality ratio to estimate under-reporting, https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html, [Online; last accessed April-2020] (Mar. 2020).
- [6] CDC COVID-19 Response Team, Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020, *MMWR Morb Mortal Wkly Rep* 69 (12) (2020) 343–346. [doi:10.15585/mmwr.mm6912e2](https://doi.org/10.15585/mmwr.mm6912e2).
- [7] J. Yang, Y. Zheng, X. Gou, K. Pu, Z. Chen, Q. Guo, R. Ji, H. Wang, Y. Wang, Y. Zhou, Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis, *Int. J. Infect. Dis.* In press. [doi:10.1016/j.ijid.2020.03.017](https://doi.org/10.1016/j.ijid.2020.03.017).
- [8] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, et al., The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, *Lancet Public Health* In press. [doi:10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6).
- [9] E. Shim, A. Tariq, W. Choi, Y. Lee, G. Chowell, Transmission potential and severity of COVID-19 in South Korea, *International Journal of Infectious Diseases* In press. [doi:10.1016/j.ijid.2020.03.031](https://doi.org/10.1016/j.ijid.2020.03.031).
- [10] A. R. Tuite, D. N. Fisman, A. L. Greer, Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada, *Can. Med. Assoc. J.* 192 (2020) 1–9. [doi:10.1503/cmaj.200476](https://doi.org/10.1503/cmaj.200476).
- [11] S. M. Kissler, C. Tedijanto, M. Lipsitch, Y. Grad, Social distancing strategies for curbing the COVID-19 epidemic, *medRxiv* [doi:10.1101/2020.03.22.20041079](https://doi.org/10.1101/2020.03.22.20041079).

- [12] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, V. Colizza, [Expected impact of lockdown in Île-de-France and possible exit strategies](#). URL https://www.epicx-lab.com/uploads/9/6/9/4/9694133/inserm-covid-19_report_lockdown_idf-20200412.pdf
- [13] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, H. Coupland, T. A. Mellan, H. Zhu, T. Berah, J. W. Eaton, P. N. P. Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, M. Baguelin, I. Blake, A. Boonyasiri, O. Boyd, L. Cattarino, C. Ciavarella, L. Cooper, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, S. van Elsland, R. F. John, H. Fu, K. Gaythorpe, L. Geidelberg, N. Grassly, W. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati-Gilani, P. Nouvellet, K. Parag, I. Siveroni, H. Thompson, R. Verity, E. Volz, C. Walters, H. Wang, Y. Wang, O. Watson, P. Winskill, X. Xi, C. Whittaker, P. G. T. Walker, A. Ghani, C. A. Donnelly, S. Riley, L. C. Okell, M. A. C. Vollmer, N. M. Ferguson, S. Bhatt, Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries [doi:10.25561/77731](https://doi.org/10.25561/77731).
- [14] D. Chen, T. Zhou, Control Efficacy on COVID-19 [arXiv:2003.00305](https://arxiv.org/abs/2003.00305).
- [15] S. Cauchemez, P.-Y. Boelle, C. A. Donnelly, N. M. Ferguson, G. Thomas, G. M. Leung, A. J. Hedley, R. M. Anderson, A.-J. Valleron, Real-time estimates in early detection of SARS, *Emerging Infectious Diseases* 12 (1) (2006) 110–113.
- [16] B. C. for Disease Control, [Covid-19 case counts and press statements](#), [Online; last accessed April-2020]. URL <http://www.bccdc.ca/health-info/diseases-conditions/covid-19-case-counts-press-statements>
- [17] British Columbia Centre for Disease Control, BC COVID-19 Data, <http://www.bccdc.ca/health-info/diseases-conditions/covid-19/data>, accessed: 2020-4-17.
- [18] D. Korzinski, S. Kurl, [COVID-19 Carelessness: Which Canadians say pandemic threat is ‘overblown’? And how are they behaving in turn?](#), [Online; last accessed April-2020] (2020). URL <http://angusreid.org/covid-19-serious-vs-overblown/>
- [19] J. F. Lawless, Statistical Models and Methods for Lifetime Data, 2nd ed., Wiley series in Probability and Statistics, Wiley, New Jersey, 2003.
- [20] A. M. Edwards, J. P. W. Robinson, J. L. Blanchard, J. K. Baum, M. J. Plank, Accounting for the bin structure of data removes bias when fitting size spectra, *Mar. Ecol. Prog. Ser.* 636 (2020) 19–33. [doi:10.3354/meps13230](https://doi.org/10.3354/meps13230).
- [21] Ministry of Citizens Services, Population Estimates – Province of British Columbia, <https://www2.gov.bc.ca/gov/content/data/statistics/people-population-community/population/population-estimates>, [Online; last accessed April-2020].
- [22] Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet Infectious Diseases* In press. [doi:10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).

- [23] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, Q. Guo, T. Song, J. He, H.-L. Yen, M. Peiris, J. Wu, SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients, *N. Engl. J. Med.* 382 (12) (2020) 1177–1179.
- [24] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia, *N. Engl. J. Med.* 382 (13) (2020) 1199–1207.
- [25] L. Tindale, M. Coombe, J. E. Stockdale, E. Garlock, W. Y. V. Lau, M. Saraswat, Y.-H. B. Lee, L. Zhang, D. Chen, J. Wallinga, C. Colijn, Transmission interval estimates suggest pre-symptomatic spread of COVID-19 (2020). [doi:10.1101/2020.03.03.20029983v1](https://doi.org/10.1101/2020.03.03.20029983v1).
- [26] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, N. Hens, Estimating the generation interval for COVID-19 based on symptom onset data (Mar. 2020).
- [27] Public Health Agency of Canada, Coronavirus disease (COVID-19): Outbreak update - Canada.ca, <https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html>, accessed: 2020-4-14 (Apr. 2020).
- [28] J. M. Hilbe, Negative Binomial Regression, Cambridge University Press, 2011.
- [29] Q. Liu, Z. Liu, J. Zhu, Y. Zhu, D. Li, Z. Gao, L. Zhou, Y. Tang, X. Zhang, J. Yang, Q. Wang, Assessing the global tendency of COVID-19 outbreak, medRxiv [doi:10.1101/2020.03.18.20038224](https://doi.org/10.1101/2020.03.18.20038224).
- [30] Stan Development Team, Prior choice recommendations, [Online; last accessed April-2020]. URL <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>
- [31] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, A. Riddell, Stan: A Probabilistic Programming Language, *J. Stat. Softw.* 76 (1). [doi:10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01).
- [32] Stan Development Team, RStan: the R interface to Stan, R package version 2.19.3.
- [33] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2019). URL <https://www.R-project.org/>
- [34] M. D. Hoffman, A. Gelman, The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learn. Res.* 15 (2014) 1593–1623.
- [35] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, 3rd Edition, Chapman & Hall, Boca Raton, FL, 2014.
- [36] J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers, *Proc. Biol. Sci.* 274 (1609) (2007) 599–604. [doi:10.1098/rspb.2006.3754](https://doi.org/10.1098/rspb.2006.3754).

- [37] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, A. C. Ghani, Report 9 - impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand, [Online; accessed April-2020]. URL <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/>
- [38] S. A. Frank, Time Delays, in: S. A. Frank (Ed.), Control Theory Tutorial: Basic Concepts Illustrated by Software Examples, Springer International Publishing, Cham, 2018, pp. 95–102.
- [39] C. I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, CMMID COVID-19 working group, P. Klepac, G. J. Rubin, W. J. Edmunds, Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK, medRxiv [doi:10.1101/2020.03.31.20049023](https://doi.org/10.1101/2020.03.31.20049023).
- [40] BlueDot: Outbreak Risk Software, <https://bluedot.global>, [Online; last accessed April-2020].
- [41] Google, Community Mobility Reports, <https://www.google.com/covid19/mobility>, [Online; last accessed April-2020].
- [42] Citymapper, Citymapper Mobility Index, <https://citymapper.com/cmi>, [Online; last accessed April-2020].
- [43] Apple, Mobility Trends Reports, <https://www.apple.com/covid19/mobility>, [Online; last accessed April-2020].
- [44] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, S. Funk, R. M. Eggo, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, Lancet Global Health 8 (4) (2020) e488–e496. [doi:10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7).
- [45] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bon-sall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, Science In press. [doi:10.1126/science.abb6936](https://doi.org/10.1126/science.abb6936).
- [46] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al., Temporal dynamics in viral shedding and transmissibility of COVID-19, medRxiv.
- [47] O. Diekmann, J. A. P. Heesterbeek, J. A. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (4) (1990) 365–382.
- [48] P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, Math. Biosci. 180 (2002) 29–48. [doi:10.1016/S0025-5564\(02\)00108-6](https://doi.org/10.1016/S0025-5564(02)00108-6).

- [49] W. A. Stein, et al., [Sage Mathematics Software \(Version 9.0\)](#), The Sage Development Team (2020).
URL <http://www.sagemath.org>
- [50] I. SageMath, [CoCalc Collaborative Computation Online](#) (2016).
URL <https://cocalc.com>

A Supplemental Methods

A.1 Computing R_0 analytically

We compute the basic reproduction number R_0 for the full model, as a function of the model parameters, using the Next-Generation method [47, 48]. We set f to be constant, dropping the subscript z , and set $u_d = 1$, $e := u_d/(u_r + u_d)$, and $u_r = (1 - e)/e$. The disease-free equilibrium is $\epsilon_0 = ((1 - e)N, 0, 0, 0, 0, eN, 0, 0, 0, 0, 0)$. We identify $E_1, E_2, I, Q, E_{1d}, E_{2d}, I_d, Q_d$ as the infection compartments, and denote the corresponding system by \mathcal{X} .

We decompose $\mathcal{X} = \mathcal{F} - \mathcal{V}$, where

$$\mathcal{F} = \begin{bmatrix} \beta(I + E_2 + f(I_d + E_{2d}))\frac{S}{N} \\ 0 \\ 0 \\ 0 \\ f\beta(I + E_2 + f(I_d + E_{2d}))\frac{S_d}{N} \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathcal{V} = \begin{bmatrix} k_1E_1 + E_1 - \frac{(1-e)}{e}E_{1d} \\ -k_1E_1 + k_2E_2 + E_2 - \frac{(1-e)}{e}E_{2d} \\ -k_2E_2 + qI + \frac{I}{D} + I - \frac{(1-e)}{e}I_d \\ -qI + \frac{Q}{D} + Q - \frac{(1-e)}{e}Q_d \\ k_1E_{1d} - E_1 + \frac{(1-e)}{e}E_{1d} \\ -k_1E_{1d} + k_2E_{2d} - E_2 + \frac{(1-e)}{e}E_{2d} \\ -k_2E_{2d} + qI_d + \frac{I_d}{D} - I + \frac{(1-e)}{e}I_d \\ -qI_d + \frac{Q_d}{D} - Q + \frac{(1-e)}{e}Q_d \end{bmatrix}.$$

Linearization around ϵ_0 gives the following Jacobian matrices for \mathcal{F} and \mathcal{V} respectively,

$$F = \begin{bmatrix} 0 & \beta(1 - e) & \beta(1 - e) & 0 & 0 & f\beta(1 - e) & f\beta(1 - e) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & f\beta e & f\beta e & 0 & 0 & f^2\beta e & f^2\beta e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} k_1 + 1 & 0 & 0 & 0 & -\frac{(1-e)}{e} & 0 & 0 & 0 \\ -k_1 & k_2 + 1 & 0 & 0 & 0 & -\frac{(1-e)}{e} & 0 & 0 \\ 0 & -k_2 & q + \frac{1}{D} + 1 & 0 & 0 & 0 & -\frac{(1-e)}{e} & 0 \\ 0 & 0 & -q & \frac{1}{D} + 1 & 0 & 0 & 0 & -\frac{(1-e)}{e} \\ -1 & & 0 & 0 & k_1 + \frac{(1-e)}{e} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -k_1 & k_2 + \frac{(1-e)}{e} & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & -k_2 & q + \frac{1}{D} + \frac{(1-e)}{e} & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & -q & \frac{1}{D} + \frac{(1-e)}{e} \end{bmatrix}.$$

We use SageMath [49] via CoCalc [50] to compute the spectral radius of FV^{-1} and simplify its expression. We find that

$$R_0 = \beta \frac{e^4(1-e)(1-f)^2k_1k_2}{(e(\frac{1}{D}+q)+1)(ek_1+1)(ek_2+1)} + \beta \frac{(ef+1-e)^2}{\frac{1}{D}+q} + \\ \beta \frac{ek_1(ef+1-e)^2}{k_2(ek_1+1)(ek_2+1)} + \beta \frac{e(ef+1-e)^2}{(ek_1+1)(ek_2+1)} + \\ \beta \frac{(ef+1-e)^2}{k_2(ek_1+1)(ek_2+1)} + \beta \frac{e^2k_1(ef^2+1-e)}{(ek_1+1)(ek_2+1)}. \quad (7)$$

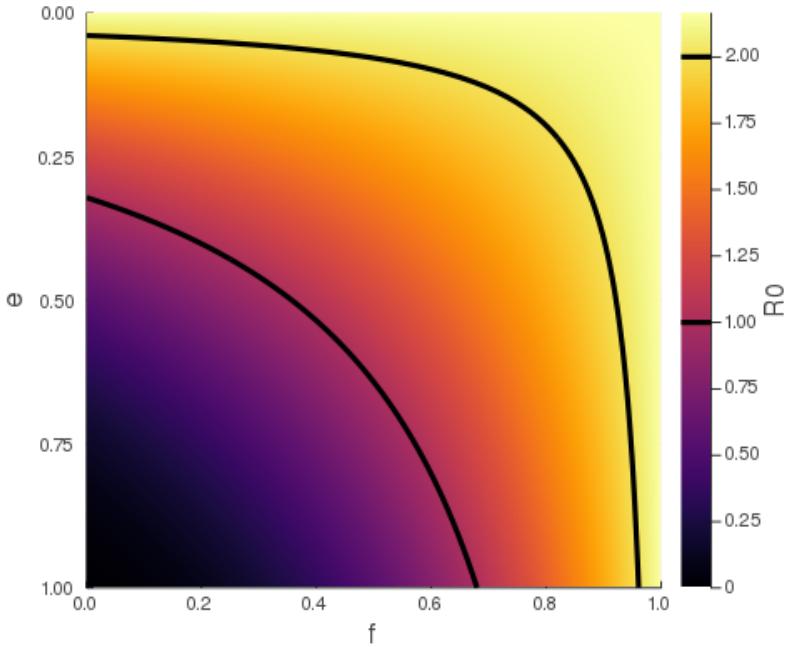


Figure S1: R_0 as a function of e and f , with fixed parameters $\beta = 0.433$, $D = 5$, $q = 0.05$, $k_1 = 0.2$, and $k_2 = 1$. The black contours represent $R_0 = 1$ and $R_0 = 2$, respectively.

Figure S1 shows an example of the relation between R_0 and the model parameters. The fixed parameters are taken from Table 1, with β computed as $R_{0b}/(D + 1/k_2)$.

A.2 Likelihood estimation of parameters for the delay distribution

We derive a likelihood function to estimate the shape and scale parameters of the Weibull delay distribution using the case-specific data that are available for some individual cases. These data are right-truncated (Figure 2), in that we are not aware of cases that will be reported in the future, yet their symptoms have already occurred in the past. This is dealt with by the likelihood function. The resulting maximum likelihood estimates are then used in the observation model to relate reported cases to model prevalence.

The data on individual cases (Figure 2) give counts h_{nr} of the number of individuals whose case was reported (test was positive) at the end of day r and whose symptoms are estimated to have started on day n . If day N is the final day of these data, then the full data set is $\{h_{nr}\}_{n=0,1,2,\dots,N; n \leq r \leq N}$, where, by definition, $h_{nr} = 0$ for $r < n$ because a case cannot be reported before the start of symptoms for these data. The counts are right-truncated on day N , since there are individuals whose symptoms started on day n who will be reported in the future (after N), but we do not yet know when. The cases are considered to be reported at the end of day r because there are values of $h_{nn} > 0$.

Assuming a Weibull distribution for the time between symptom onset and reporting of a case, then the probability of a case whose symptoms start on day n being reported on day $r \geq n$ is

$$p'_{nr} = \int_r^{r+1} w(\tau' - n; \lambda, k) d\tau', \quad (8)$$

$$= \int_{r-n}^{r-n+1} w(\tau; \lambda, k) d\tau, \quad (9)$$

where the probability density function for the Weibull distribution with shape $k > 0$ and scale $\lambda > 0$ is

$$w(\tau; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{\tau}{\lambda}\right)^{k-1} e^{-(\tau/\lambda)^k}, & \tau \geq 0, \\ 0, & \tau < 0. \end{cases} \quad (10)$$

The integral in (9) is needed because the Weibull is a continuous function of time (and time is continuous in the differential equation model). We define the cumulative distribution function as

$$W(x; \lambda, k) = P(X \leq x; \lambda, k) = \int_0^x w(\tau; \lambda, k) d\tau. \quad (11)$$

Then

$$p'_{nr} = W(r - n + 1) - W(r - n), \quad (12)$$

where we have dropped the explicit λ and k in $W(\cdot)$ for clarity.

Equation (9) holds for $r > N$, but the data are only available up to $r = N$. Consider that today is day $N = 5$, then we have a matrix of p'_{nr} values:

$$(p'_{nr}) = \left(\begin{array}{cccccc|cccccc} p'_{00} & p'_{01} & p'_{02} & p'_{03} & p'_{04} & p'_{05} & | & p'_{06} & p'_{07} & \dots \\ 0 & p'_{11} & p'_{12} & p'_{13} & p'_{14} & p'_{15} & | & p'_{16} & p'_{17} & \dots \\ 0 & 0 & p'_{22} & p'_{23} & p'_{24} & p'_{25} & | & p'_{26} & p'_{27} & \dots \\ 0 & 0 & 0 & p'_{33} & p'_{34} & p'_{35} & | & p'_{36} & p'_{37} & \dots \\ 0 & 0 & 0 & 0 & p'_{44} & p'_{45} & | & p'_{46} & p'_{47} & \dots \\ 0 & 0 & 0 & 0 & 0 & p'_{55} & | & p'_{56} & p'_{57} & \dots \end{array} \right). \quad (13)$$

The probabilities in each row sum to 1 (since every case will eventually become reported, else it would not become a ‘case’). The cases to the right of the vertical dashed line are the ones that we do not yet have data on; for example, people whose symptoms started on day 1 and whose case will be reported tomorrow (day 6, i.e. p'_{16}).

We need to calculate p_{nr} , the probability that a case that is reported on day r exhibited the onset of symptoms on day n . This requires taking the values in the above matrix that are to the left of the dashed line, and normalising each column such that it sums to 1, i.e.

$$p_{nr} = \frac{p'_{nr}}{\sum_{i=0}^r p'_{ir}}, \quad n, r \leq N. \quad (14)$$

The denominator is

$$\begin{aligned} \sum_{i=0}^r p'_{ir} &= \sum_{i=0}^r [W(r-i+1) - W(r-i)], \\ &= W(r+1) - W(0), \\ &= W(r+1), \end{aligned} \quad (15)$$

such that

$$p_{nr} = \frac{W(r-n+1) - W(r-n)}{W(r+1)}, \quad n, r \leq N. \quad (16)$$

For $N = 5$ we have the matrix of p'_{nr} values:

$$(p_{nr}) = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} & p_{05} \\ 0 & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ 0 & 0 & p_{22} & p_{23} & p_{24} & p_{25} \\ 0 & 0 & 0 & p_{33} & p_{34} & p_{35} \\ 0 & 0 & 0 & 0 & p_{44} & p_{45} \\ 0 & 0 & 0 & 0 & 0 & p_{55} \end{pmatrix}, \quad (17)$$

which is an upper triangular matrix of size 6×6 , i.e. $(N+1) \times (N+1)$. With an extra day of data (setting $N = 6$) the matrix has an extra row (people with onset of symptoms on day 6 can now be reported) and an extra column (the reported values on day 6):

$$(p_{nr}) = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} & p_{05} & p_{06} \\ 0 & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \\ 0 & 0 & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\ 0 & 0 & 0 & p_{33} & p_{34} & p_{35} & p_{36} \\ 0 & 0 & 0 & 0 & p_{44} & p_{45} & p_{46} \\ 0 & 0 & 0 & 0 & 0 & p_{55} & p_{56} \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{66} \end{pmatrix}. \quad (18)$$

Given the data h_{nr} and these probabilities, we develop a multinomial log-likelihood function [19], adapting the approach of [20], to estimate the parameters λ and k . The log-likelihood func-

tion for λ and k , given the counts $\{h_{nr}\}$, is

$$l(\lambda, k | \{h_{nr}\}) = \log \left[\prod_{n=0}^N \prod_{r=n}^N p_{nr}^{h_{nr}} \right], \quad (19)$$

$$= \sum_{n=0}^N \sum_{r=n}^N h_{nr} \log p_{nr}, \quad (20)$$

$$= \sum_{n=0}^N \sum_{r=n}^N h_{nr} \left\{ \log \left[W(r-n+1) - W(r-n) \right] - \log W(r+1) \right\}, \quad (21)$$

where the $W(\cdot)$ terms depend on λ and k . This is maximised numerically to give maximum likelihood estimates λ_{MLE} and k_{MLE} that are used in the differential equation model.

A key point to realise is that the definitions of p_{nr} for $r = 0$ to 5 (columns 1–6) are the same in (17) and (18), but their values will change from (17) to (18) because fitting the extra data from day 6 will update λ_{MLE} and k_{MLE} , which affects all the estimated values of the probabilities. Also, not all of the daily reported cases used in the differential equation model have details regarding onset of symptoms, i.e. $\sum_{n=0}^N h_{nr} \leq C_r$. Using the available data we obtain estimates (with 95% univariate confidence intervals) of $\lambda_{MLE} = 9.85$ (9.30 – 10.46) and $k_{MLE} = 1.73$ (1.60 – 1.86), as shown in Figure S2. Using the maximum likelihood estimates, the mean delay is 8.78 days.

In the differential equation model, time is continuous and defined such that C_r , the number of recorded cases on day r , is equal to the number of cases recorded by the end of day r . The number of people per day who become symptomatic throughout day n is the total number moving from the E_2 and E_{2d} compartments to the I and I_d compartments, which is

$$\int_{n-1}^n k_2 [E_2(\tau) + E_{2d}(\tau)] d\tau. \quad (22)$$

The expected number of cases that are reported on day r is then

$$\widehat{C}_r = \psi_r \sum_{n=1}^r \int_{n-1}^n k_2 [E_2(\tau) + E_{2d}(\tau)] w(r-\tau) d\tau, \quad (23)$$

$$= \psi_r \int_0^r k_2 [E_2(\tau) + E_{2d}(\tau)] w(r-\tau) d\tau, \quad (24)$$

where ψ_r represents the sampling fraction on day r and $w(\cdot)$ represents the Weibull distribution with parameters λ_{MLE} and k_{MLE} . If $\psi_r = 1$ then all estimated infectious people are tested and then become reported cases; $\psi_r < 1$ represents a reduction in expected cases on day r due to not everyone being tested. By changing variables $\tau = r - s$, this becomes

$$\widehat{C}_r = -\psi_r \int_r^0 k_2 [E_2(r-s) + E_{2d}(r-s)] w(s) ds, \quad (25)$$

$$= \psi_r \int_0^r k_2 [E_2(r-s) + E_{2d}(r-s)] w(s) ds. \quad (26)$$

The differential equation model is started from a time earlier than $r = 0$, so the integration here can go back further because the model calculates, for example, $E_2(-1)$; $r < 0$ does not occur in

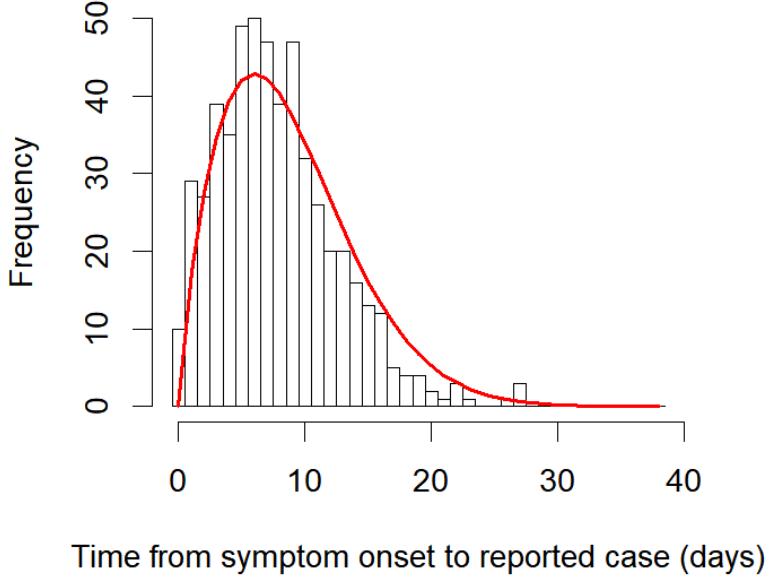


Figure S2: Histogram of times between onset of symptoms and a case being reported (from the case-specific data). The red line shows the maximum likelihood fit of the Weibull distribution, taking into account the right-truncation of the data. The distribution estimates longer delays (mean of 8.78 days) than shown in the data (mean of 7.79 days) due to the fit accounting for right-truncation—there has not yet been the opportunity to record many long delays (Figure 2).

the above likelihood calculations because, by definition, there are no data for such times. This requires setting some maximum delay between symptom onset and a case being reported that is large enough for $w(s)$ to be negligible. Based on Figure S2 we use a maximum delay of 45 days (the 99.99992% quantile), giving

$$\widehat{C}_r = \psi_r \int_0^{45} k_2 \left[E_2(r-s) + E_{2d}(r-s) \right] w(s) ds, \quad (27)$$

which is calculated and compared to the data (C_r) via the negative binomial observation model.

A.3 Model Validation

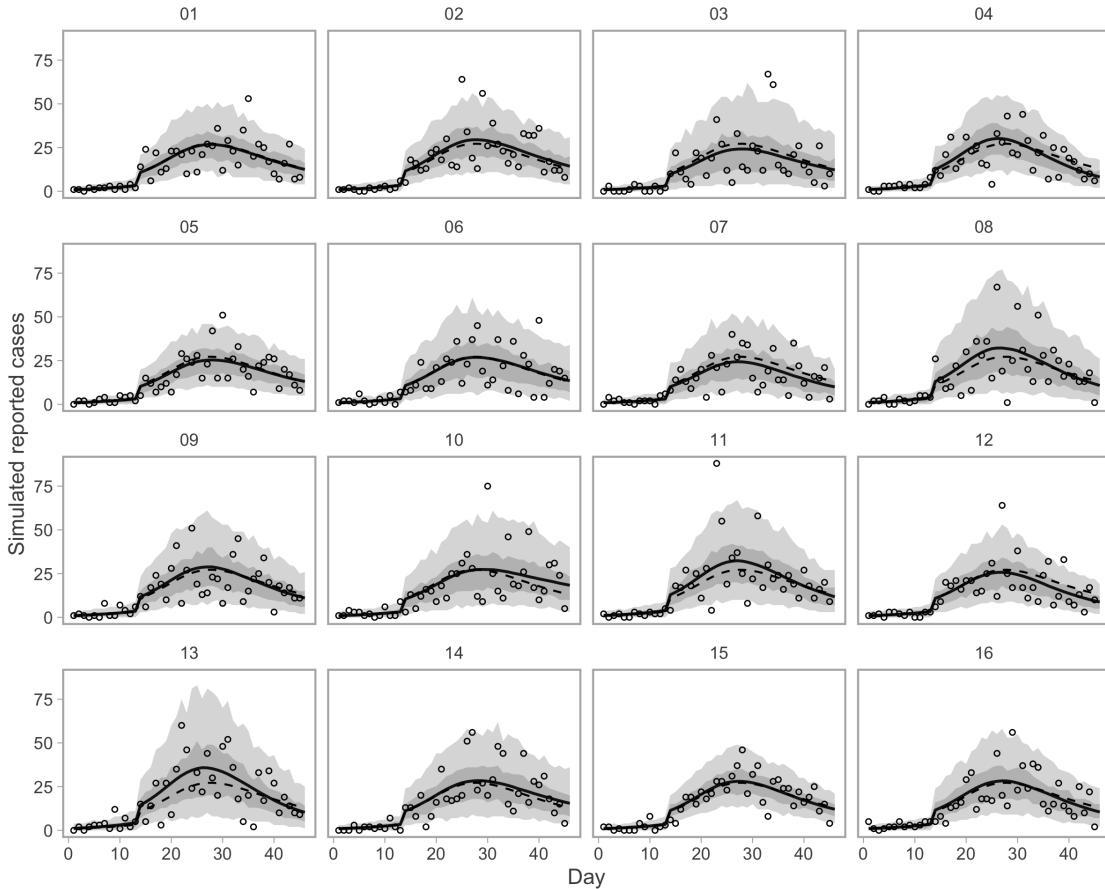


Figure S3: Simulation testing the ability to recover time series of reported cases. Dashed lines represent the true simulated time series. Solid lines represent means of the posterior. Dots represent observed simulated data. Shaded ribbons represent 50% and 90% credible intervals on new observations and should ideally encompass 50% and 90% of the simulated data points, respectively.

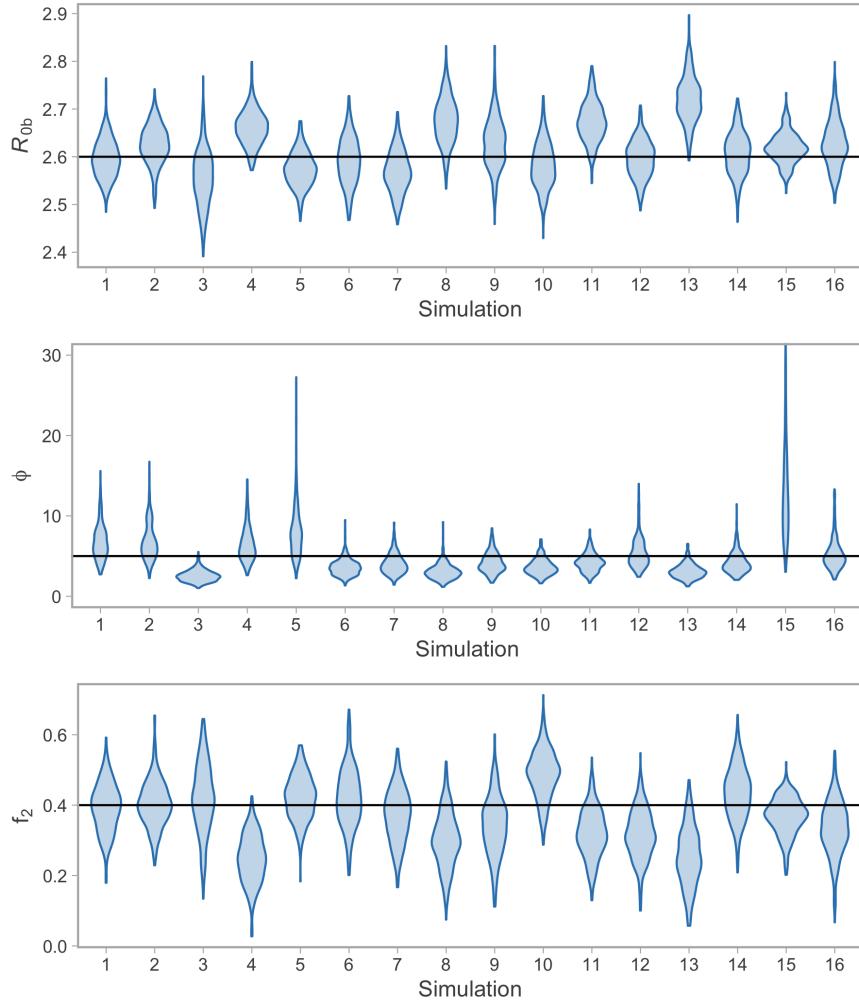


Figure S4: Simulation testing the ability to recover parameter values. The parameter R_{0b} is the basic reproductive number of the model without interventions. The parameter ϕ is the dispersion parameter of the negative binomial observation model. The fraction of contacts that are removed is $1 - f_2$. The “violins” illustrate the posterior density distribution across 16 simulation examples. The true values used in the simulation are indicated by the horizontal black lines.

B Supplemental Results

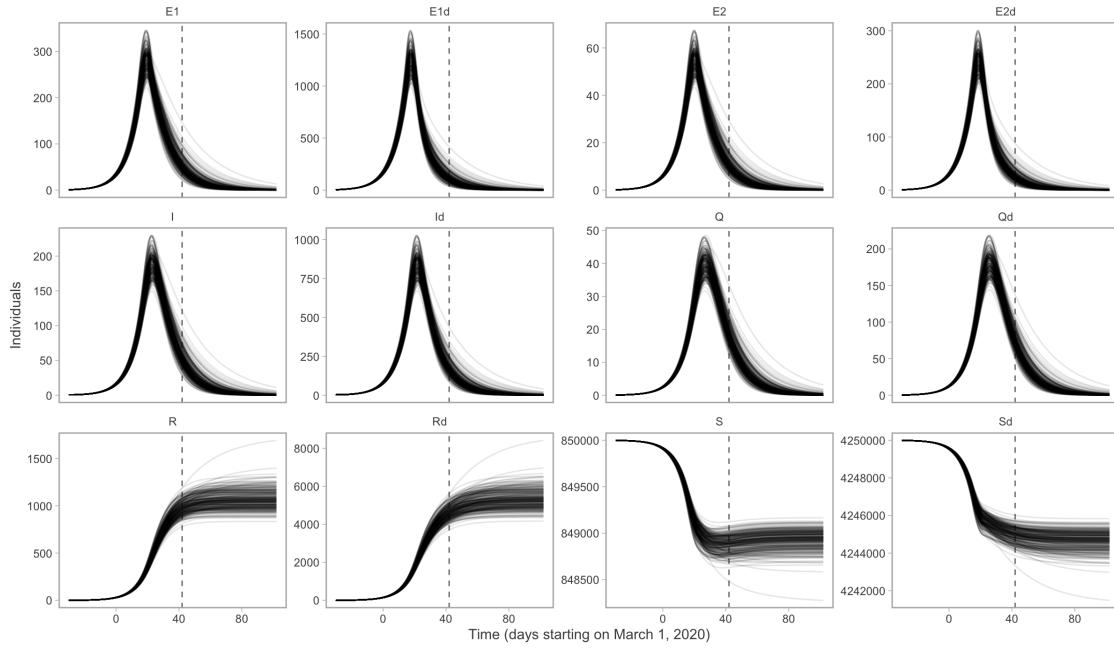


Figure S5: Posterior distributions of the underlying state variables. The time-dependent variables are the numbers of individuals who are: susceptible (S), exposed to the virus, asymptomatic and not infectious (E_1), exposed, asymptomatic and infectious (E_2), infectious (I), quarantined (Q), and recovered and deceased (R). Recovered individuals are assumed to be immune. There are analogous variables for individuals practising physical distancing, represented by subscript d, i.e., S_d , E_{1d} , E_{2d} , I_d , Q_d , and R_d .

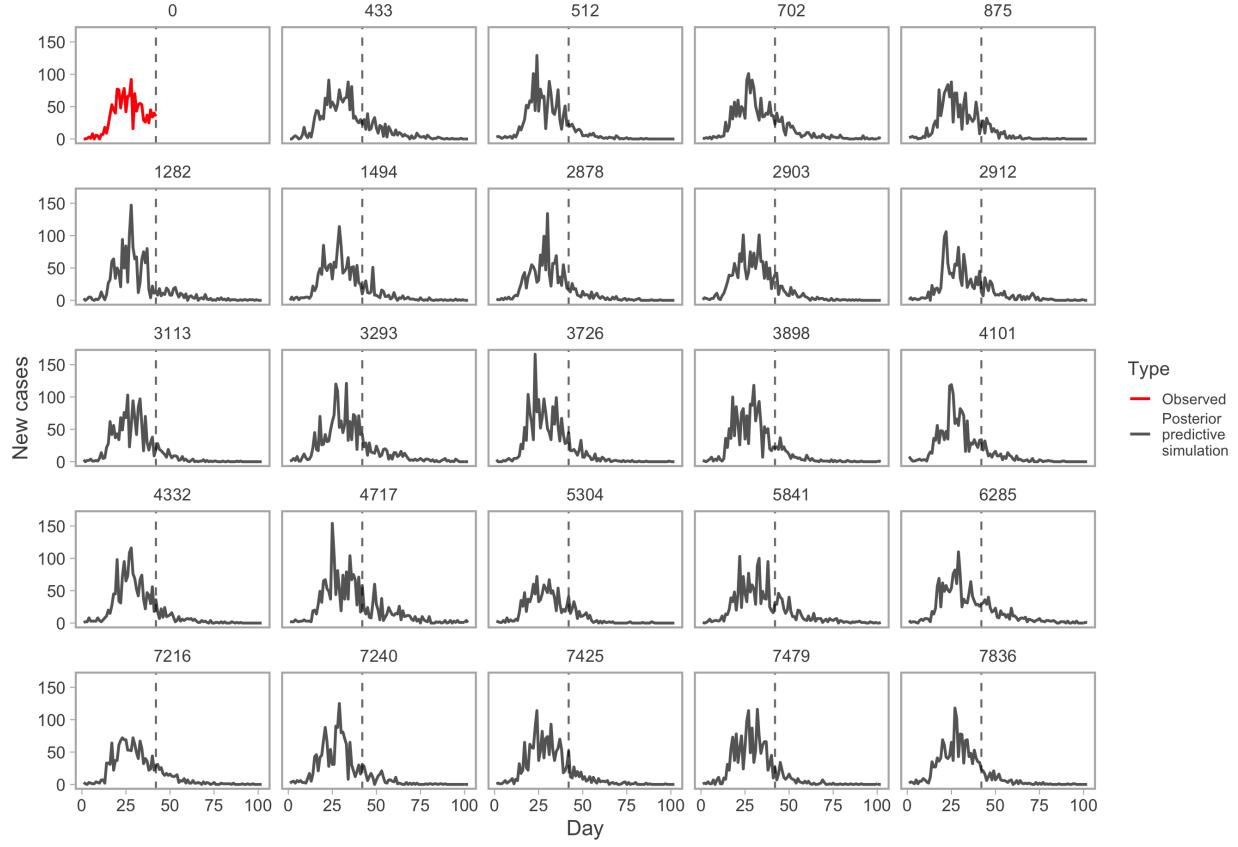


Figure S6: Example posterior-predictive replicates from the model fit. The first panel (red line) represents the observed data. All other panels represent example draws from the posterior-predictive distribution. Numbers above panels represent randomly selected Markov chain Monte Carlo (MCMC) iterations. This plot helps evaluate whether the observed data are consistent with data generated by the model.

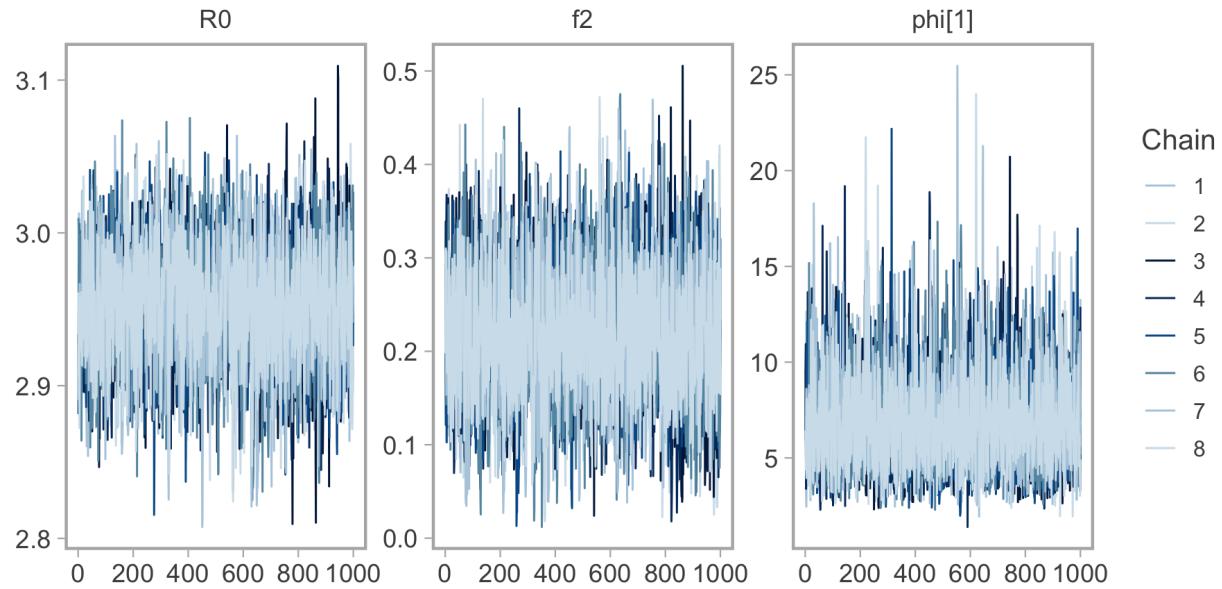


Figure S7: Trace plots of MCMC (Markov chain Monte Carlo) samples from parameter distributions to assess chain convergence. The panel labelled “R0” actually refers to R_{0b} . The panel labelled “phi[1]” refers to the dispersion parameter ϕ from the negative binomial observation model.

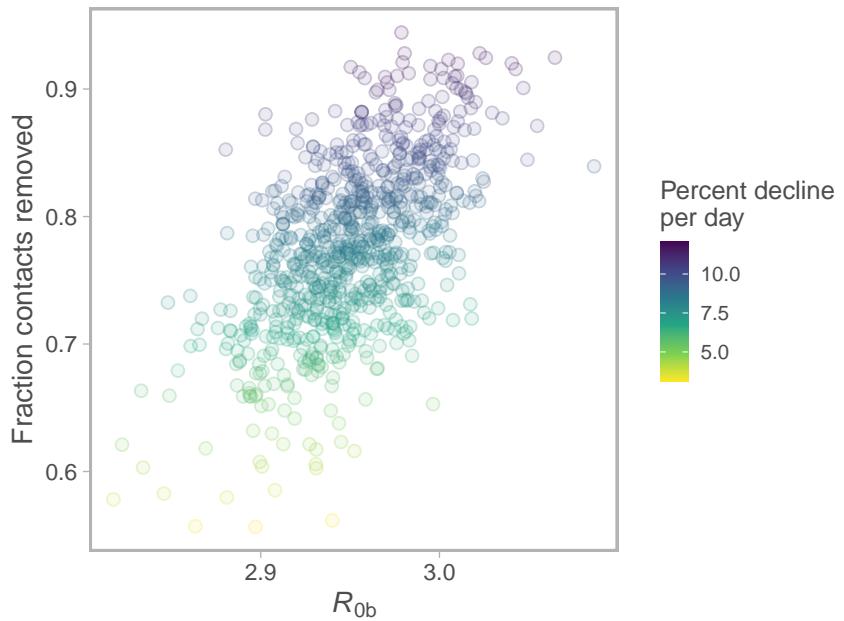


Figure S8: Joint posterior of fraction contacts removed vs. R_{0b} . Colour represents percent decline per day over the last 30 days of the projection. Dots represent individual draws from the joint posterior.

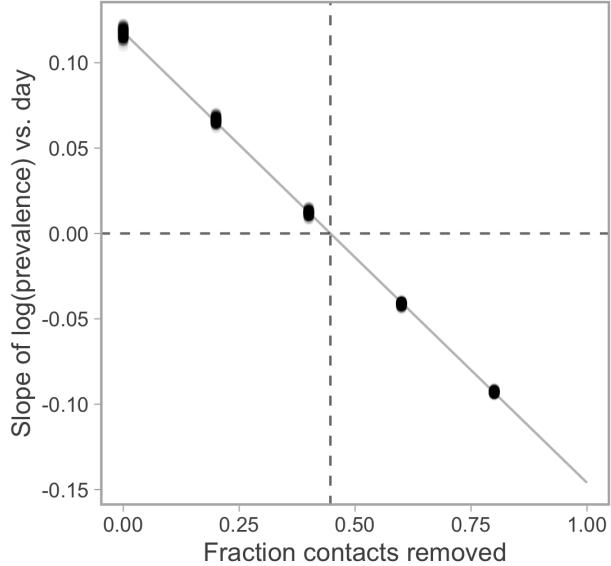


Figure S9: Linear regression of [the slope of log(prevalence) over days (for the last 30 days of the projection)] vs. [fraction of contacts removed]. Vertical dashed line indicates the critical value below which the rate of change of prevalence becomes positive. Dots represent individual posterior draws.

C Sensitivity Analysis

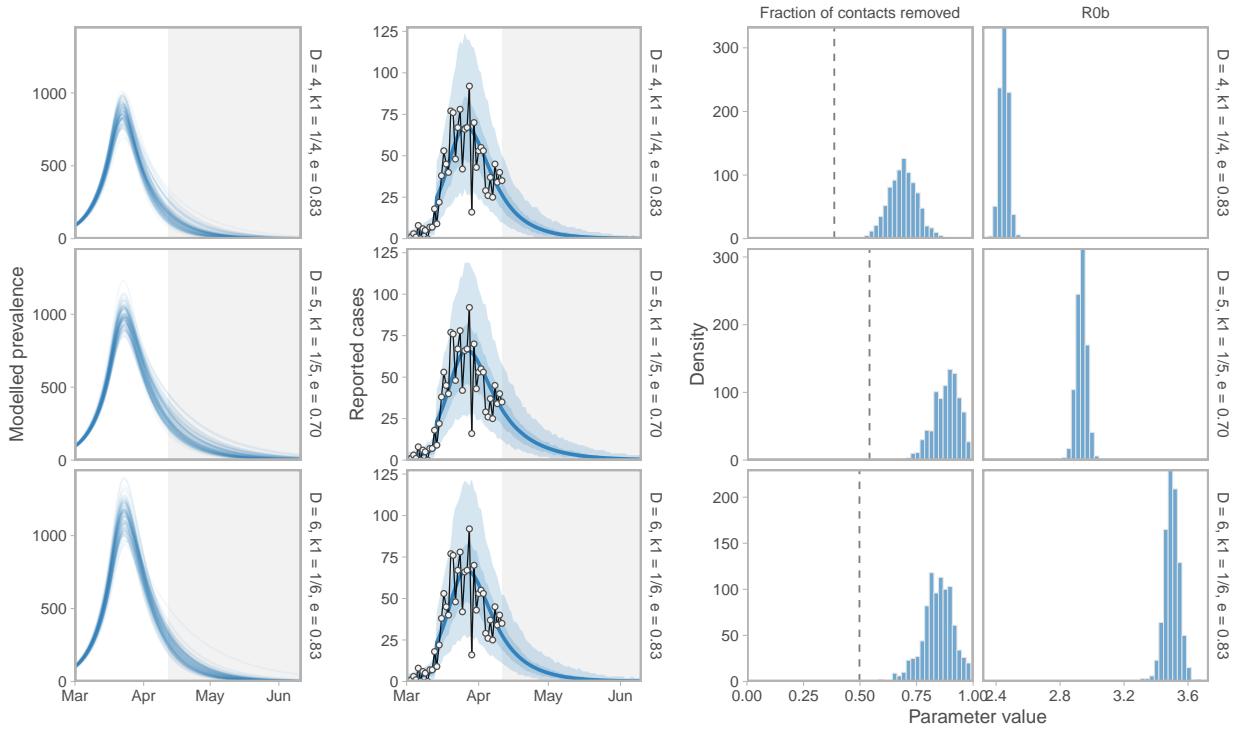


Figure S10: Sensitivity of prevalence and case-count projections, R_{0b} , and the fraction-of-contacts-removed estimate ($1 - f_2$) to three sets of alternative parameter values: (1) $D = 4, k_1 = 1/4, u_r = 0.2$ (shorter duration); (2) $D = 5, k_1 = 1/5, u_r = 0.3/0.7$ (a lower proportion of people physical distancing: 70% vs. 83% in the main analysis); and (3) $D = 6, k_1 = 1/6, u_r = 0.2$ (longer duration). Histograms illustrate the posterior distribution of MCMC samples. Dashed vertical line represents the threshold value below which an exponential increase occurs.

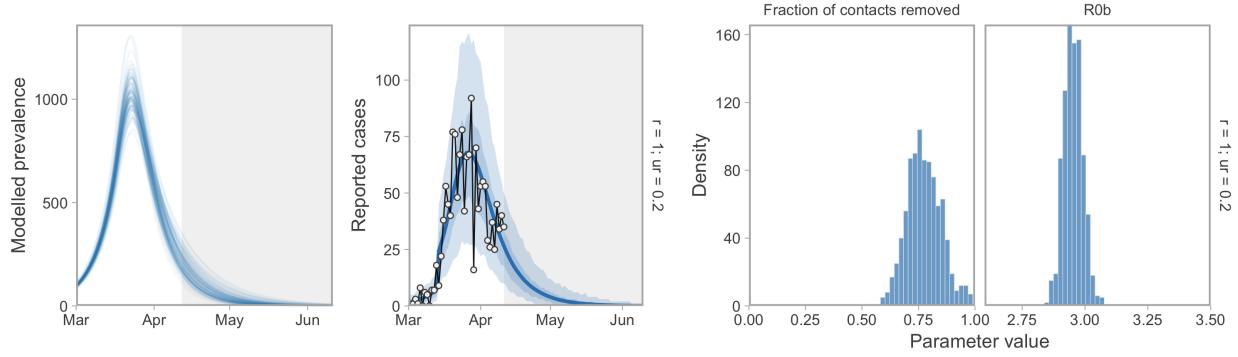


Figure S11: Prevalence and case-count projections, R_{0b} , and the fraction-of-contacts-removed estimate ($1 - f_2$) with distancing rate parameters $u_d = 1$ and $u_r = 0.2$. The ratio e is the same as at baseline but these rates are 10 times higher. The model dynamics are sensitive to e but not to the choice of rates. Accordingly, these estimates are as in Figure 4.

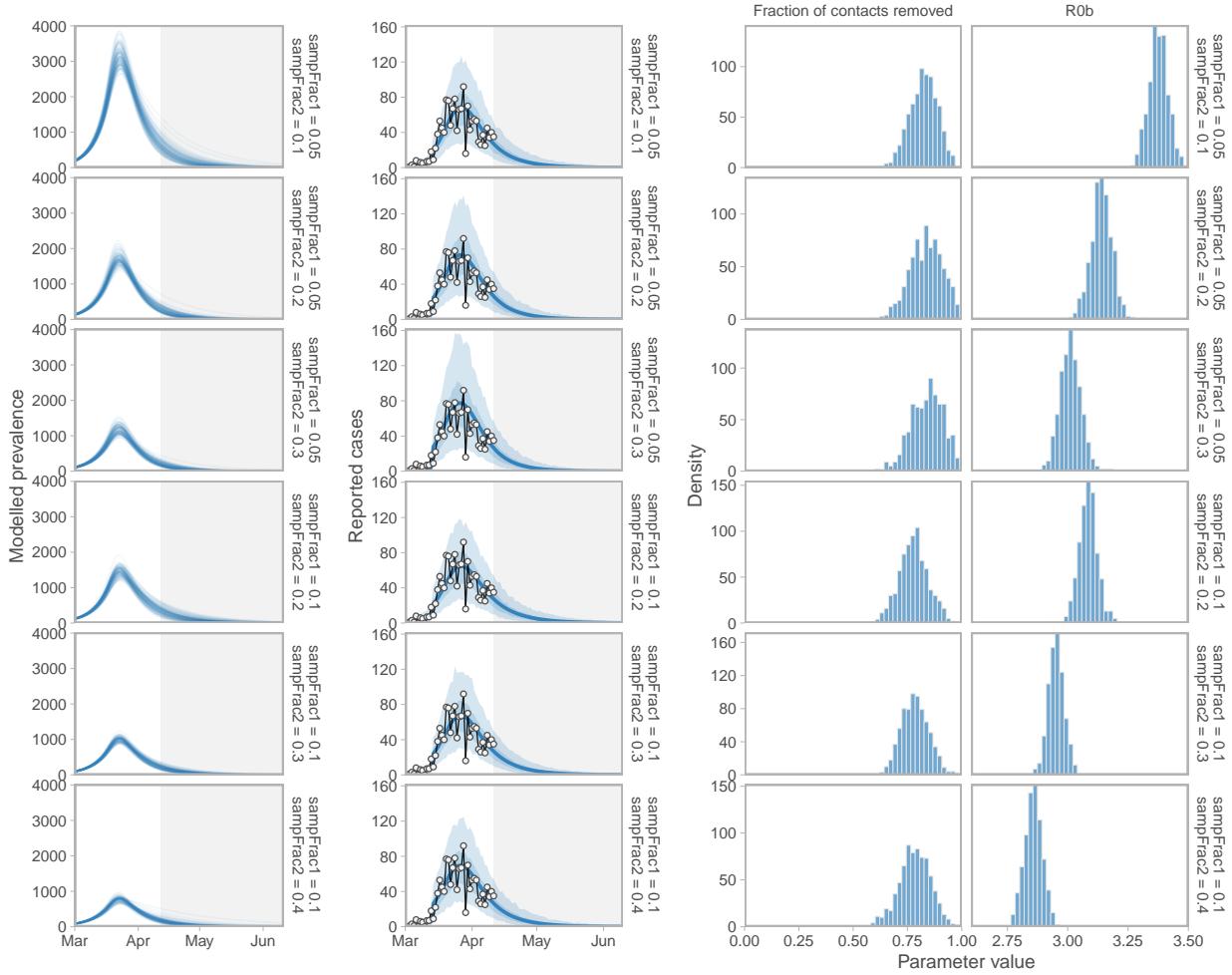


Figure S12: Sensitivity of prevalence and case-count projections, R_{0b} , and the fraction that contacts are removed ($1 - f_2$) to assumed fractions of positive cases sampled. “sampFrac1” refers to the assumed sample fraction before March 14, 2020 and “sampFrac2” refers to the assumed sample fraction on and after March 14, 2020. Histograms illustrate the posterior distribution of MCMC samples.

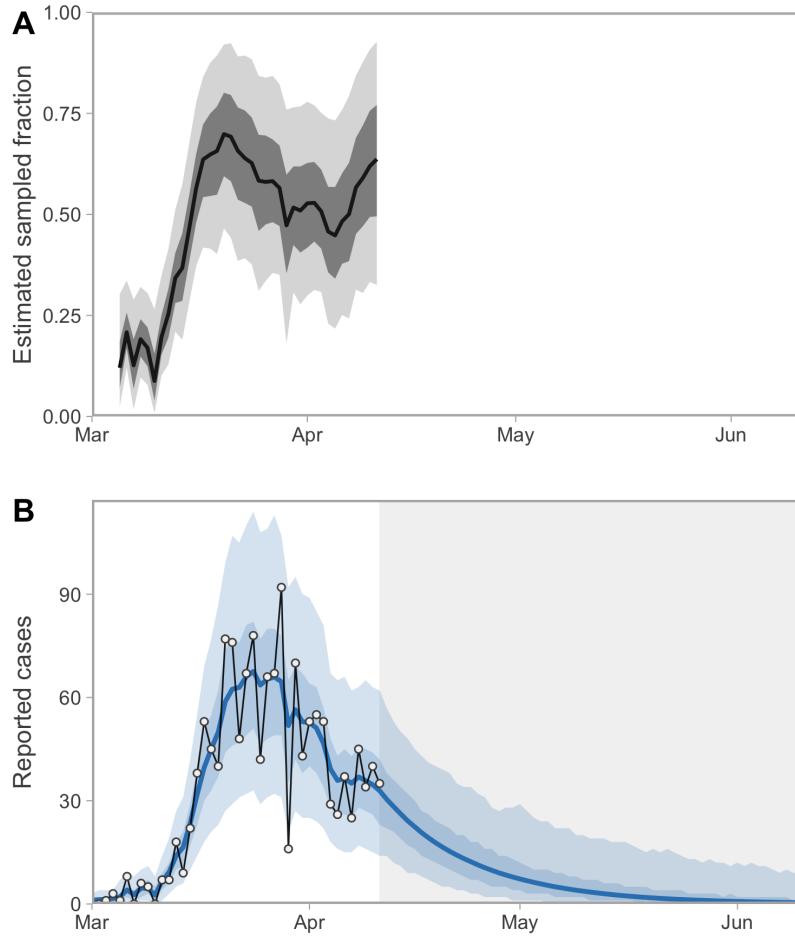


Figure S13: Alternative model fitted with a random walk on the fraction of positive cases sampled. (A) Estimated sampled fraction random walk. The random walk starts on the fifth day (it is fixed at 0.1 before then) and then evolves with a fixed standard deviation of 0.1. We placed a Beta prior on the initial value with a mean and standard deviation of 0.2 and 0.2. (B) Resulting model fit with a sampled-fraction random walk. Shaded ribbons represent 50% and 90% credible intervals. Thick lines represent medians (A) and means (B) of the posterior distribution. Although the shape in panel A is plausible, we deemed the scale of the estimated values to be implausibly high and therefore used a fixed value of 0.3 after day 14 in the main models in this paper.