Quantifying dynamics of SARS-CoV-2 transmission suggests that epidemic control and avoidance is feasible through instantaneous digital contact tracing: Supplementary Information

Luca Ferretti*, Chris Wymant*, Michelle Kendall, Lele Zhao,
Anel Nurtay, David Bonsall, Christophe Fraser†

*Contributed equally
†To whom correspondence should be addressed: christophe.fraser@bdi.ox.ac.uk

Big Data Institute, Li Ka Shing Centre for Health Information and Discovery,
University of Oxford, Oxford, UK

March 6, 2020

Supplementary Text

Inference of the distribution of generation times

We collected data on 40 source-recipient pairs with known onset of symptoms from public sources. This set of transmission events spans different countries (see Supplementary Figure 2) and social relations (households, family members, colleagues, acquaintances, random contacts). For each transmission event $1 \to 2$, we include the dates of onset of symptoms $t_{s,1}$ and $t_{s,2}$, the intervals of exposure $[e_{1,L}, e_{1,R}]$ and $[e_{2,L}, e_{2,R}]$, and the reporting date T_r. The intervals of exposure must satisfy the conditions $e_{1,R} \leq \min(t_{s,1}, e_{2,R})$, $e_{2,L} \geq e_{1,L}$ and $e_{2,R} \leq t_{s,2}$.

First, we discretize the incubation/generation time distributions

$$I(j) = \int_{j-0.5}^{j+0.5} dt\ i(t) \quad , \quad \Omega(j|\Theta_\omega) = \int_{j-0.5}^{j+0.5} dt\ \omega(t) \tag{1}$$

where the incubation time distribution $i(t)$ is taken from [Lauer et al 2020], and Θ_ω denotes the parameters of the generation time distribution $\omega(t)$. Then, for a given transmission event $1 \to 2$, we define the likelihood of observing the transmission and the subsequent time of onset of symptoms in the recipient as

$$\mathcal{L}_{\text{trans}}(\Theta_\omega) = P[1 \to 2, t_{s,2}|t_{s,1}, \Theta_\omega] \tag{2}$$

which can be obtained by summing over all possible infection times $t_{i,1}$ and $t_{i,2}$, resulting in the explicit form

$$\mathcal{L}_{\text{trans}}(\Theta_\omega) \propto \sum_{t_{i,1}=e_{1,L}}^{e_{1,R}} e^{-r(T_r - t_{i,1})} I(t_{s,1} - t_{i,1}) \sum_{t_{i,2}=e_{2,L}}^{e_{2,R}} \Omega(t_{i,2} - t_{i,1}|\Theta_\omega) I(t_{s,2} - t_{i,2}) \tag{3}$$

up to a multiplicative term independent of $\omega(t)$ which has no effect on likelihood maximisation. The growth rate r corresponds to a doubling time of 5 days for exponentially growing epidemics (China except Hong Kong, Italy, South Korea) while it is assumed to be 0 for countries where local transmission is limited.

For simplicity, we ignore correlations between transmission events. We define the composite likelihood across all 40 pairs (i.e. the approximate global likelihood that neglects correlations) as the product of the likelihoods $\mathcal{L}_{\text{trans}}(\Theta_\omega)$ of all pairs. We then test multiple functional forms for $\omega(t)$ (lognormal, gamma, Weibull), inferring their parameters $\hat{\Theta}_\omega$ by Maximum Composite Likelihood. Finally, we select the functional form based on the lowest AIC. Confidence intervals for the parameters of the distribution are based on likelihood ratios, while confidence intervals for the values of the curve are based on likelihood profiling.

This approach relies not only on the timing of transmissions and onset of symptoms, but it also assumes that transmission pairs themselves are chosen at random from the epidemic. This is not generally the case. To test the robustness of our approach, we relax this assumption by considering also the probability conditional on the occurrence of the transmission $1 \to 2$ at any time:

$$\mathcal{L}_{|\text{trans}}(\Theta_\omega) = P[t_{s,2}|1 \to 2, t_{s,1}, \Theta_\omega] = \frac{P[1 \to 2, t_{s,2}|t_{s,1}, \Theta_\omega]}{P[1 \to 2|t_{s,1}, \Theta_\omega]} = \frac{\mathcal{L}_{\text{trans}}(\Theta_\omega)}{\sum_{t'_{2,s}} P[1 \to 2, t'_{2,s}|t_{s,1}, \Theta_\omega]} \quad (4)$$

This approach does not extract any information from the fact that a transmission actually occurred. The MCL inference uses only the timing of exposure and onset of symptoms. The correspondence between the two approaches is therefore a good test of robustness.

Posterior probability of pre-symptomatic transmission

For each transmission event, the likelihood $\mathcal{L}_{\text{trans}}(\Theta_\omega)$ described above can be decomposed as

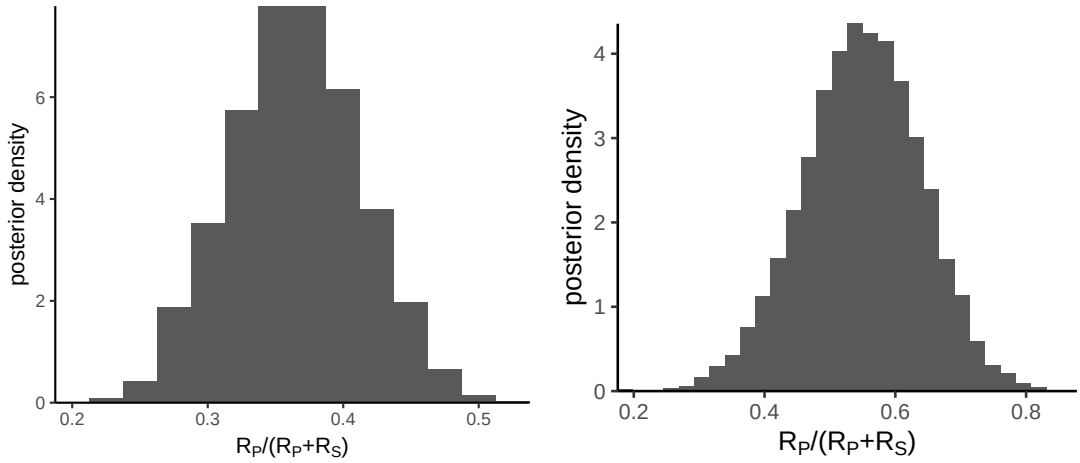
$$\mathcal{L}_{\text{trans}} = \mathcal{L}_{\text{pre}} + \mathcal{L}_{\text{pre}} \quad (5)$$

where the pre-symptomatic term $\mathcal{L}_{\text{pre}}$ includes only the cases with $t_{i,2} < t_{s,1}$ and the symptomatic term $\mathcal{L}_{\text{sym}}$ includes all other cases. We assume a prior probability $p_{\text{pre}} = p_{\text{sym}} = 0.5$ for pre-symptomatic and symptomatic transmission. Then, for each transmission event, we can obtain the posterior probability of pre-symptomatic transmission via a Bayesian approach:

$$P[\text{presymptomatic transmission}] = \frac{p_{\text{pre}}\mathcal{L}_{\text{pre}}(\hat{\Theta}_\omega)}{p_{\text{pre}}\mathcal{L}_{\text{pre}}(\hat{\Theta}_\omega) + p_{\text{sym}}\mathcal{L}_{\text{sym}}(\hat{\Theta}_\omega)} \quad (6)$$

where all likelihoods are evaluated at the parameters $\hat{\Theta}_\omega$ that maximise the composite likelihood.

The distribution of the fraction of transmissions that occurred before onset of symptoms can be estimated by assigning each event as pre-symptomatic or symptomatic at random according to its posterior probability. The empirical distribution of this quantity is obtained from 10,000 random extractions from the posterior, shown in the left panel of Supplementary Figure 1.



Supplementary Figure 1: Uncertainty on the contribution of pre-symptomatic transmissions to all transmissions from infected individuals that will eventually show symptoms. Left: posterior distribution from the 40 transmission pairs analysed in this study. Right: posterior distribution from the infectiousness model approach.

$\beta(\tau)$ and the renewal equation

In an epidemic which is growing exponentially, in a deterministic manner, driven by human-to-human[1] transmission, the incidence $I(t)$ can be described by the renewal equation:

$$I(t) = \int_0^\infty I(t-\tau)\beta(\tau)d\tau, \tag{7}$$

In words, Equation 7 says that the incidence now is set by the rate at which people were infected at all previous times, weighted by how infectious those people are now. $\beta(\tau)$ is the mean rate at which an individual infects others a time τ after being infected itself. Here we take $\beta(\tau)$ to be independent of the stage of the epidemic (calendar time t): we neglect depletion of susceptible individuals through acquired immunity, changing contact patterns etc. over the timescale of the data informing our estimations of $\beta(\tau)$. After $\beta(\tau)$ has been determined, we may consider how to change it through interventions to reduce infectiousness. If one's direct and indirect infectiousness is certain to be zero after having been infected for a time T say, we only need consider the previous time window T of the epidemic – replacing the upper limit of the integral in Equation 7 by T for convenience. We take T to be infinite for generality, with $\beta(\tau)$ tending to zero at large times. Substituting into Equation 7 an exponentially growing incidence, $I(t) = I_0 e^{rt}$, gives the condition

$$1 = \int_0^\infty e^{-r\tau}\beta(\tau)d\tau, \tag{8}$$

$\beta(\tau)$ can be written as the product of two things: R_0 and the unit-normalised function $w(\tau)$

$$\beta(\tau) = R_0 w(\tau), \quad \text{with} \quad R_0 = \int_0^\infty \beta(\tau)d\tau \tag{9}$$

[1]Equation 7 describes only human-to-human transmission (though this may be indirectly via the environment, as we clarify). Excluding vector-borne diseases, zoonosis events do not scale with the number of people currently infected, and so become a negligible contribution after human-to-human transmission has begun driving exponential growth.

$w(\tau)$ is the generation time distribution – the probability density function for the time between an individual becoming infected and their subsequent onward transmission events. R_0 is the basic reproduction number. If the exponential growth rate r and the generation time distribution $w(\tau)$ have been estimated, R_0 is determined by Equation 8, i.e.

$$R_0 = 1/\int_0^\infty e^{-r\tau} w(\tau) d\tau, \tag{10}$$

We decompose $\beta(\tau)$ without any loss of generality into the following distinct contributions:

- Direct transmissions from asymptomatic individuals – those who never develop symptoms. The degree to which individuals show symptoms is of course a continuum, but a threshold can be defined for clinical purposes (i.e. sub-clinical and clinical infections) or for epidemiological purposes. We define P_a as the proportion of such individuals among all infected individuals, and $\beta_a(\tau)$ as their mean infectiousness at age-of-infection τ.

- Direct transmissions from pre-symptomatic individuals (currently without symptoms, but who will develop symptoms later). We define $\beta_p(\tau)$ as the mean infectiousness of these individuals at age-of-infection τ, conditional upon their being pre-symptomatic, which has probability $1 - s(\tau)$ where $s(\tau)$ is the cumulative distribution function of the incubation period distribution.

- Direct transmissions from symptomatic individuals (including those who have stopped showing symptoms, in general, if infectiousness may outlast symptoms), with infectiousness $\beta_s(\tau)$ conditional on having started symptoms.

- Indirect transmission via the environment. We define $\beta_e(\tau)$ as the mean rate of contaminating one's environment (with the mean being over all asymptomatic, pre-symptomatic and symptomatic individuals at age-of-infection τ). Let $E(l)$ be the rate at which contaminated environment infects new individuals a time lag l after having been contaminated. The environmentally mediated infectiousness of an individual infected a time τ ago is given by the total effect of their previous environmental contamination now: $\int_{l=0}^{\tau} \beta_e(\tau - l) E(l) d\tau$.

We therefore have, in general,

$$\beta(\tau) = P_a\beta_a(\tau) + (1 - P_a)(1 - s(\tau))\beta_p(\tau) + (1 - P_a)s(\tau)\beta_s(\tau) + \int_{l=0}^{\tau} \beta_e(\tau - l)E(l)dl \tag{11}$$

Integrating each of these terms separately gives their respective contribution to R_0:

$$R_0 = R_a + R_p + R_s + R_e \tag{12}$$

We make the following simplifying assumptions about the contributions described above, compared to the general case.

- Asymptomatic individuals are assumed to have an infectiousness proportional to that of symptomatic individuals: $\beta_a(\tau) = x_a\beta_s(\tau)$.

- Pre-symptomatic individuals are assumed to have an infectiousness equal to that of symptomatic individuals at the same age of infection: $\beta_p(\tau) = \beta_s(\tau)$

- The rate at which individuals contaminate their environment, $\beta_e(\tau)$ is assumed to be proportional to the direct infectiousness of symptomatic individuals $\beta_s(\tau)$. The proportionality constant can be absorbed into the function $E(l)$ which multiplies $\beta_e(\tau)$, so that we have $\beta_e(\tau) = \beta_s(\tau)$.

The $\beta(\tau)$ we consider is therefore

$$\beta(\tau) = \underbrace{P_a x_a \beta_s(\tau)}_{\text{asymptomatic}} + \underbrace{(1-P_a)(1-s(\tau))\beta_s(\tau)}_{\text{pre-symptomatic}} + \underbrace{(1-P_a)s(\tau)\beta_s(\tau)}_{\text{symptomatic}} + \underbrace{\int_{l=0}^{\tau} \beta_s(\tau-l)E(l)dl}_{\text{environmental}} \quad (13)$$

Derivation of the impact of interventions

To compute the impact of interventions on R_0, we follow the approach and notation of [Fraser et al 2004]. One difference in notation is that we use $s(\tau)$ to denote the probability of having started showing symptoms, corresponding to Fraser et al.'s $(1 - S(\tau))$.

Denote the efficacy of contact tracing by ϵ_T and the efficacy of isolation by ϵ_I. Denote by $Y(t, \tau, \tau')$ the number of individuals at time t who were infected at a time $t - \tau$ by individuals who were in turn infected at a time $t - \tau'$.

The generalised Kermack-McKendrick equations (referred to as Von Foerster equations) for $Y(t, \tau, \tau')$ are

$$\frac{\partial Y(t,\tau,\tau')}{\partial t} + \frac{\partial Y(t,\tau,\tau')}{\partial \tau} + \frac{\partial Y(t,\tau,\tau')}{\partial \tau'} = 0 \quad (14)$$

$$Y(t,0,\tau) = \beta(\tau)\left[1-\epsilon_I s(\tau)\right] \int_{\tau}^{\infty} \left[1 - \epsilon_T + \epsilon_T \frac{1-s(\tau')}{1-s(\tau'-\tau)}\right] Y(t,\tau,\tau')d\tau' \quad (15)$$

where the first equation (14) corresponds to translational invariance, i.e. $Y(t+dt, \tau+dt, \tau'+dt) = Y(t, \tau, \tau')$.

For an epidemic that grows exponentially at rate r, the dominant solution to equation (14) should satisfy the ansatz

$$Y(t,\tau,\tau') = y(\tau'-\tau)e^{r(t-\tau)} \quad (16)$$

which results in the next generation equation for y:

$$y(\tau) = e^{-r\tau}\beta(\tau)\left[1-\epsilon_I s(\tau)\right] \int_{0}^{\infty} \left[1 - \epsilon_T + \epsilon_T \frac{1-s(\rho+\tau)}{1-s(\rho)}\right] y(\rho)d\rho \quad (17)$$

Hence, the growth rate after the interventions correspond to the value of r for which the functional linear 'next-generation' operator

$$\mathcal{N}_r y = e^{-r\tau}\beta(\tau)\left[1-\epsilon_I s(\tau)\right] \int_{0}^{\infty} \left[1 - \epsilon_T + \epsilon_T \frac{1-s(\rho+\tau)}{1-s(\rho)}\right] y(\rho)d\rho \quad (18)$$

has the largest eigenvalue equal to 1. In other words, given the eigenvalue equation $\mathcal{N}_r y = \nu_r^{max} y$, r is determined by $\nu_r^{max} = 1$, in a functional generalization of the Euler-Lotka equation.

If R_0 is unknown, but the generation time distribution $\omega(\tau) = \beta(\tau)/R_0$ is known, then we consider the operator

$$\mathcal{O} = \mathcal{N}_0/R_0 = \omega(\tau)\left[1-\epsilon_I s(\tau)\right] \int_{0}^{\infty} \left[1 - \epsilon_T + \epsilon_T \frac{1-s(\rho+\tau)}{1-s(\rho)}\right] (\cdot)d\rho \quad (19)$$

By construction, this operator has maximal eigenvalue $1/R_0$ if and only if the value of R_0 corresponds to a growth rate $r = 0$ after interventions. Then, the inverse of the largest eigenvalue of $\mathcal{O}$ is precisely the maximum value of R_0 for which $R \leq 1$ in the presence of interventions.

This approach includes potential transmission from a fraction P_a of completely asymptomatic individuals. Denote by x_a the relative infectiousness of those individuals (which could be different from 1). For the purpose of this model, this is effectively equivalent to having $P_a x_a$

infectious individuals with relative infectiousness 1. Hence, the effective fraction of asymptomatic individuals is $P_a x_a/(1 - P_a + P_a x_a)$. Then, the fraction of infected individuals that have already shown symptoms before time τ post infection corresponds to

$$s(\tau) = \frac{1 - P_a}{1 - P_a + P_a x_a} \int_0^\tau i(\tau')d\tau' \tag{20}$$

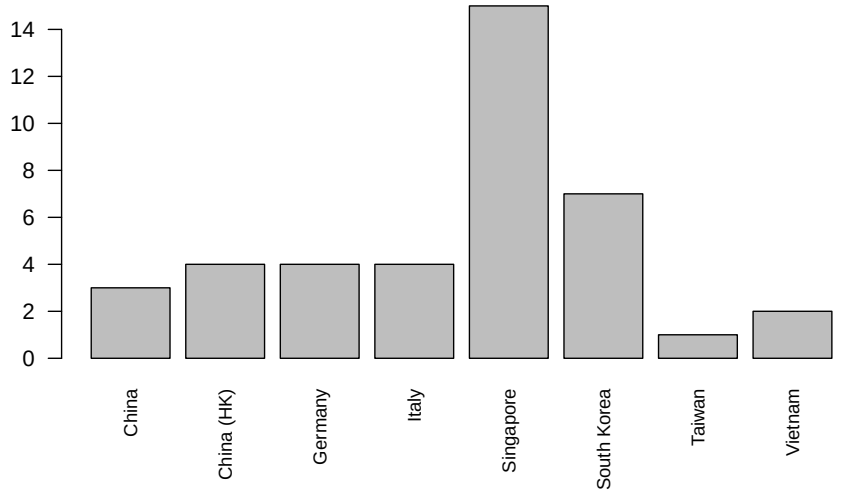
where $i(\tau)$ is the incubation time distribution for individuals that will eventually become symptomatic.

This approach can be modified to include environmental transmission as well. As a first approximation, neglecting the delays due to the persistence of SARS-CoV-2 in the environment, environmental transmission events can be treated as cases of untraceable transmission. Isolation and quarantine are effective in preventing environmental spread, since they stop the infected individual from releasing the virus in the environment. (The only relevant exception is environment transmission within isolation facilities, which is not prevented by isolation - quite the opposite.) We assume here that all environmental transmission events correspond simply to untraceable transmissions, and that the rate of viral shedding in the environment is proportional to the person-to-person infectiousness. Then, the theory developed above can be applied by replacing the efficacy of contact tracing $\epsilon_T \to \epsilon_T(1 - R_e/R_0)$, where the factor $1 - R_e/R_0$ corresponds to the fraction of traceable transmissions and ϵ_T is the efficacy per traceable transmission. Hence, the results above can be generalised to include environmental transmission by considering the next-generation operator

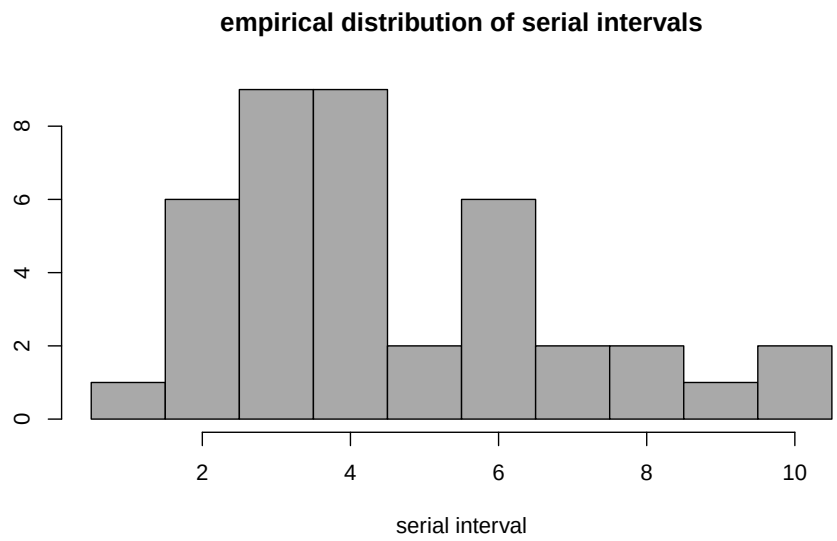
$$\mathcal{N}_r = e^{-r\tau}\beta(\tau)\left[1 - \epsilon_I s(\tau)\right] \int_0^\infty \left[1 - \epsilon_T \left(1 - \frac{R_e}{R_0}\right) \frac{s(\rho) - s(\rho + \tau)}{1 - s(\rho)}\right](\cdot)d\rho \tag{21}$$

Numerically, the eigenfunctions corresponding to the largest eigenvalues can be found by iterating the operator $\mathcal{N}_r$. Up to a constant factor, the exact solution is given by $\lim_{k\to\infty} \mathcal{N}_r^k y_0$ where y_0 is an arbitrary initial condition; in practice, $\mathcal{N}_r^k y_0$ for large enough k provides a good approximation to the solution, with exponentially fast convergence (at a rate given by the ratio of second- and first-most dominant eigenvalues per application of the operator). We found that fewer than 10 iterations provided an excellent approximation in the parameter space and for the distributions and values of R_0 considered in this paper.

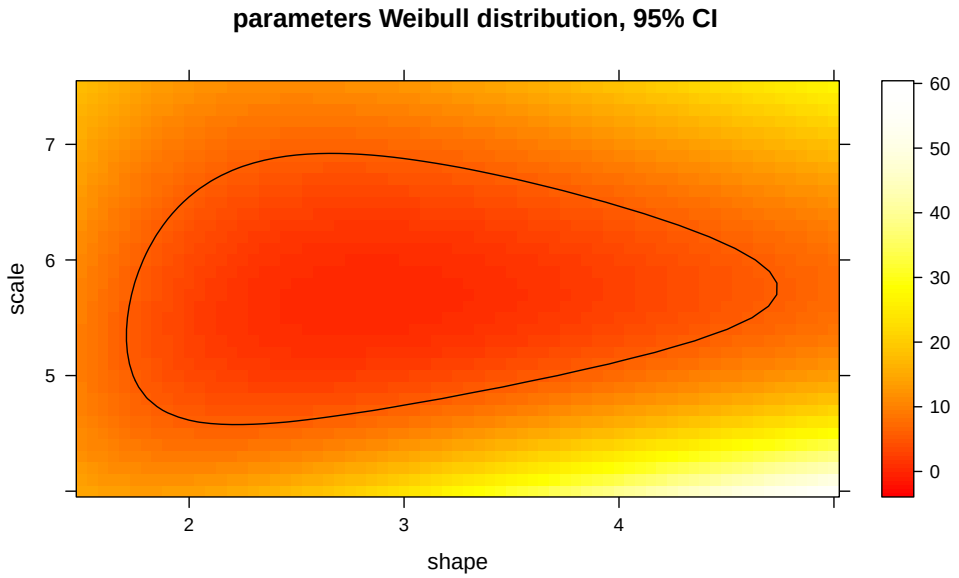
Supplementary Figures



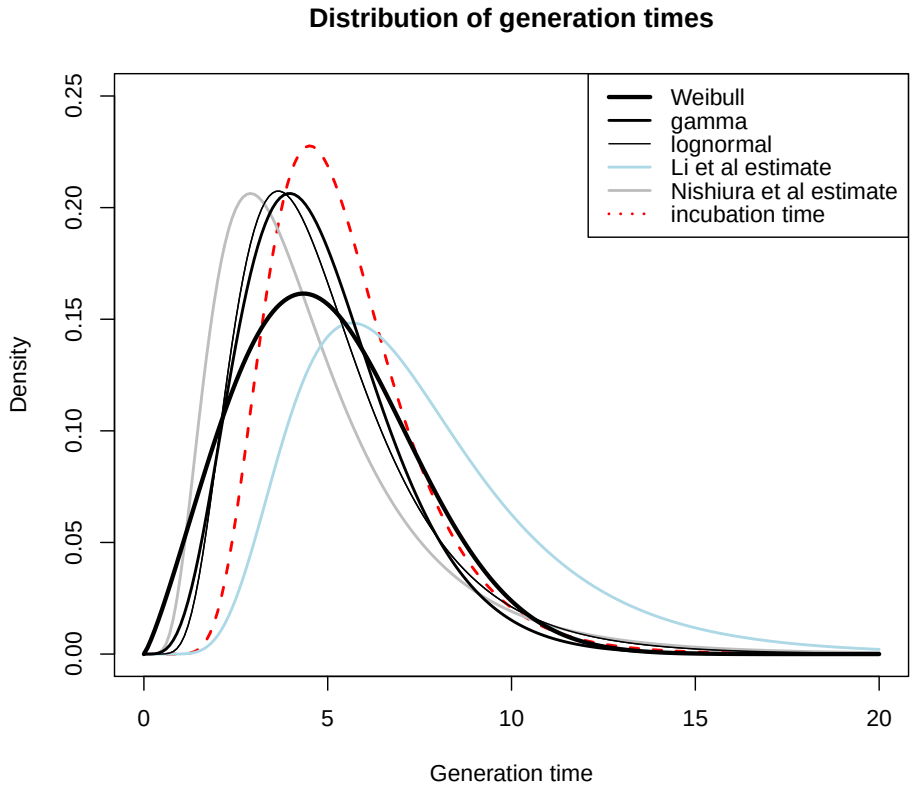
Supplementary Figure 2: Geographical distribution of transmission occurrence among the 40 source-recipient pairs analysed in this study.



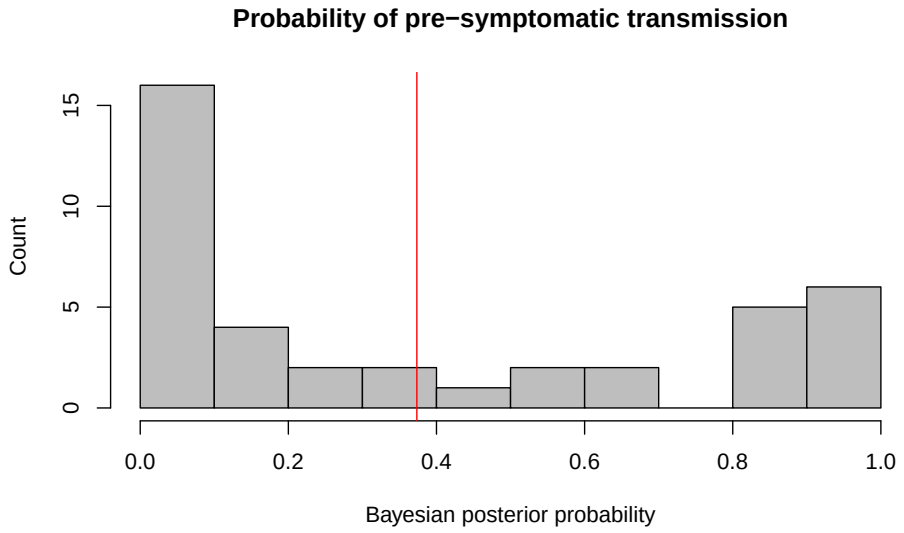
Supplementary Figure 3: Distribution of serial intervals (i.e. time from onset of symptoms in source to onset of symptoms in recipient) from the 40 transmission pairs analysed in this study.



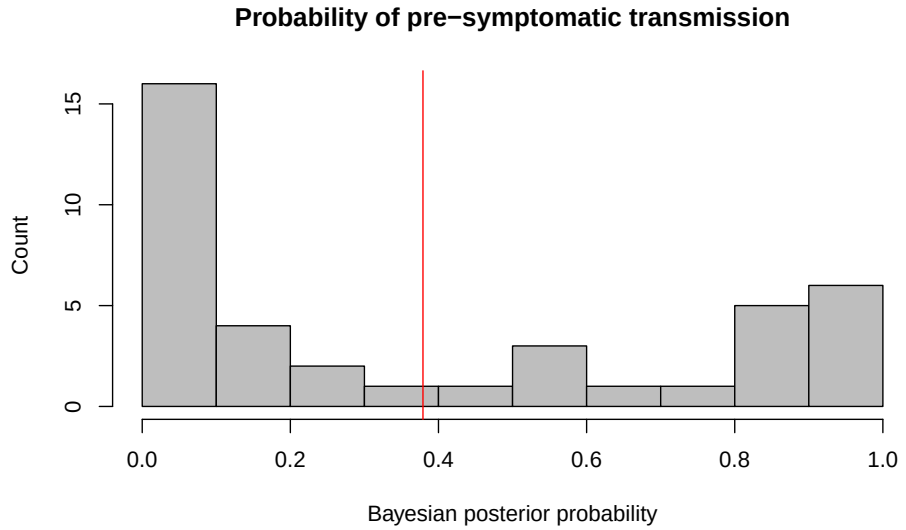
Supplementary Figure 4: Contour of the 95% confidence interval for the shape and scale parameters of the Weibull distribution of generation times. The colours represent the $-2\Delta \log \mathcal{L}$ test statistics (lower values correspond to more likely parameters).



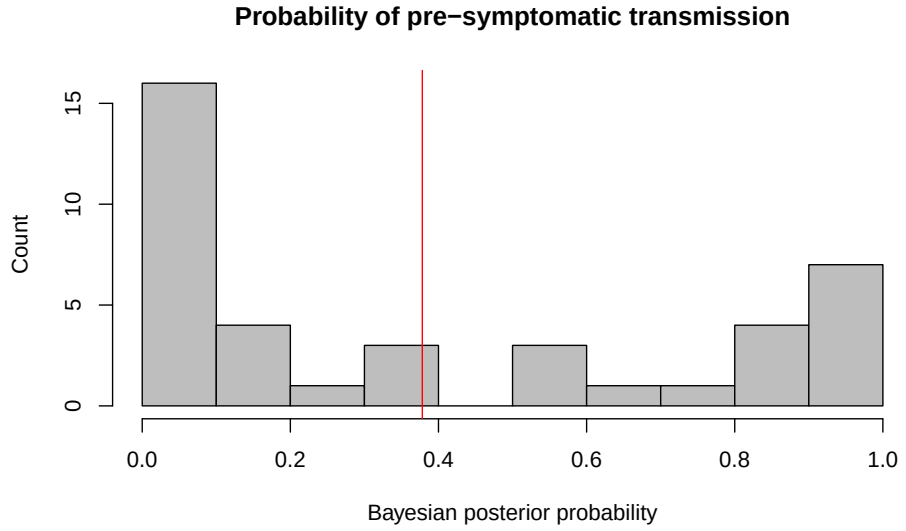
Supplementary Figure 5: Same as Figure 2 in Main Text, but the generation time distribution is inferred from the likelihood function conditional on the occurrence of the transmission events.



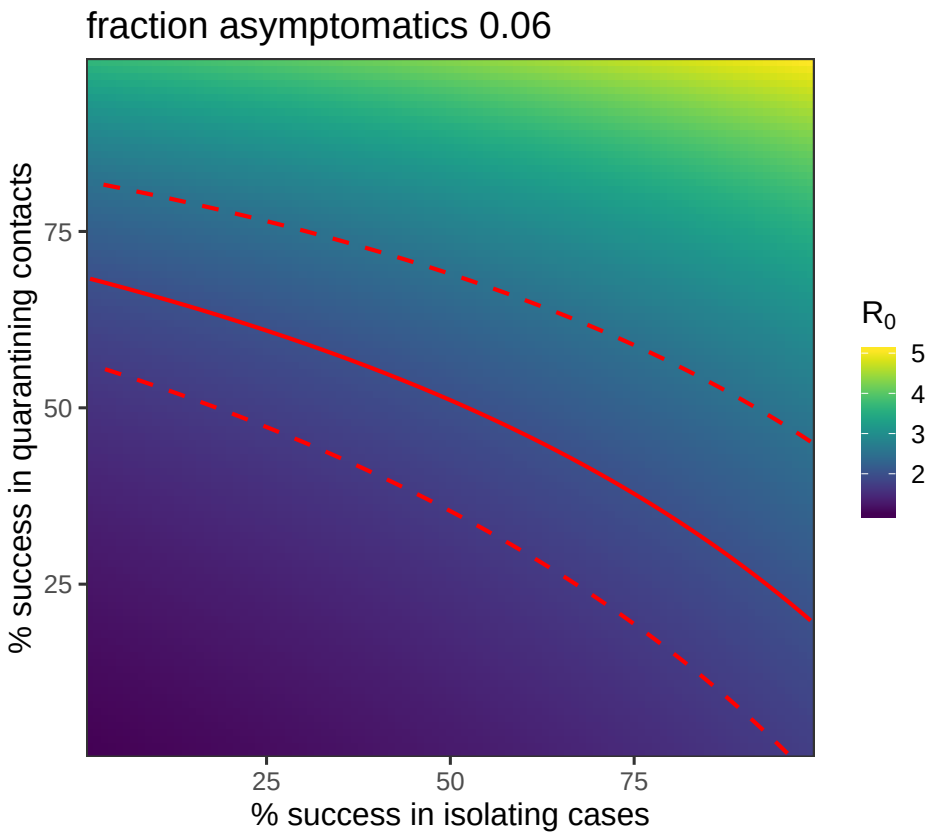
Supplementary Figure 6: Same as Figure 3 in Main Text, but assuming a gamma distribution of generation times.



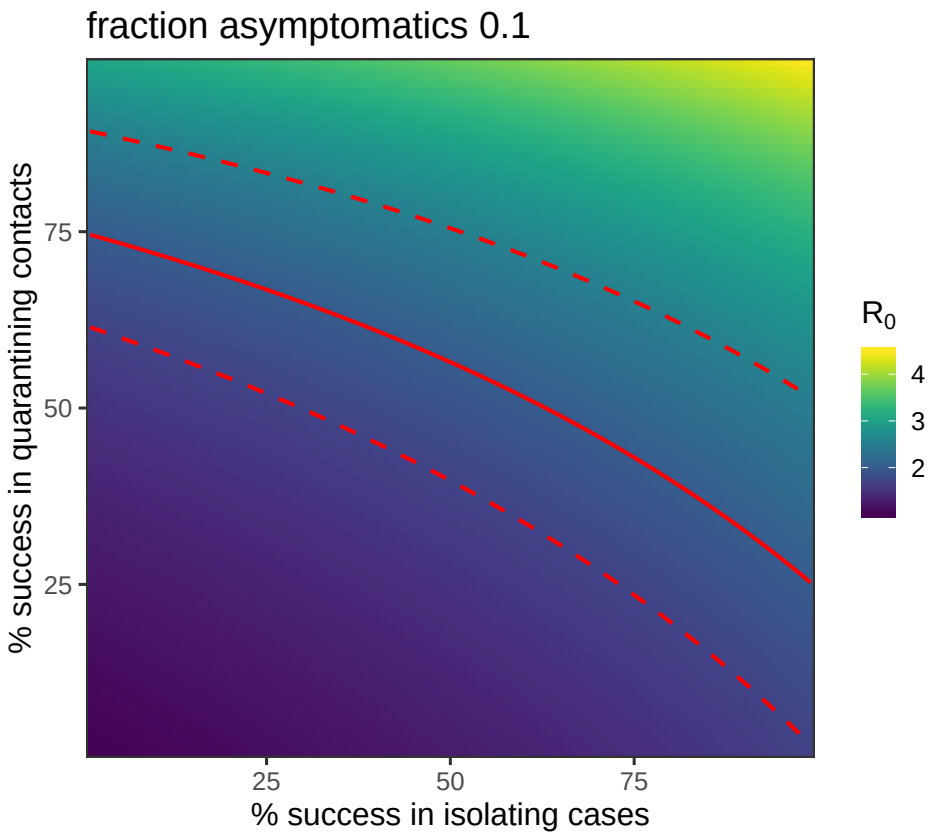
Supplementary Figure 7: Same as Figure 3 in Main Text, but assuming a gamma distribution of generation times.



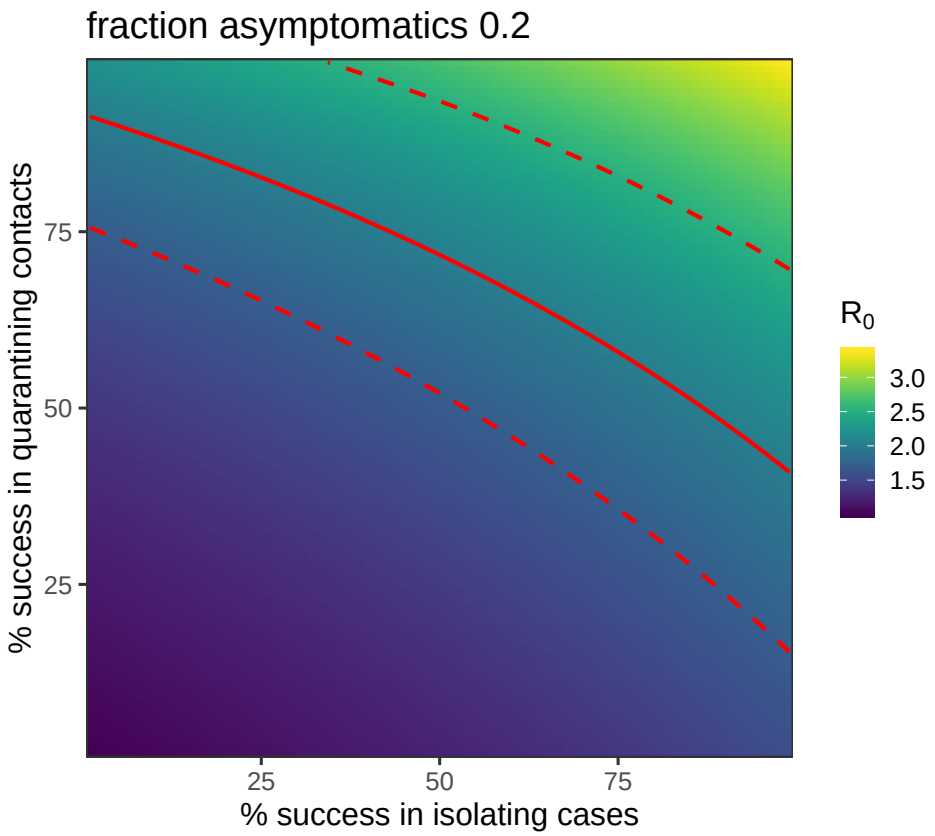
Supplementary Figure 8: Same as Figure 3 in Main Text, but the posterior probabilities of presymptomatic transmission are inferred from the likelihood function conditional on the occurrence of the transmission events.



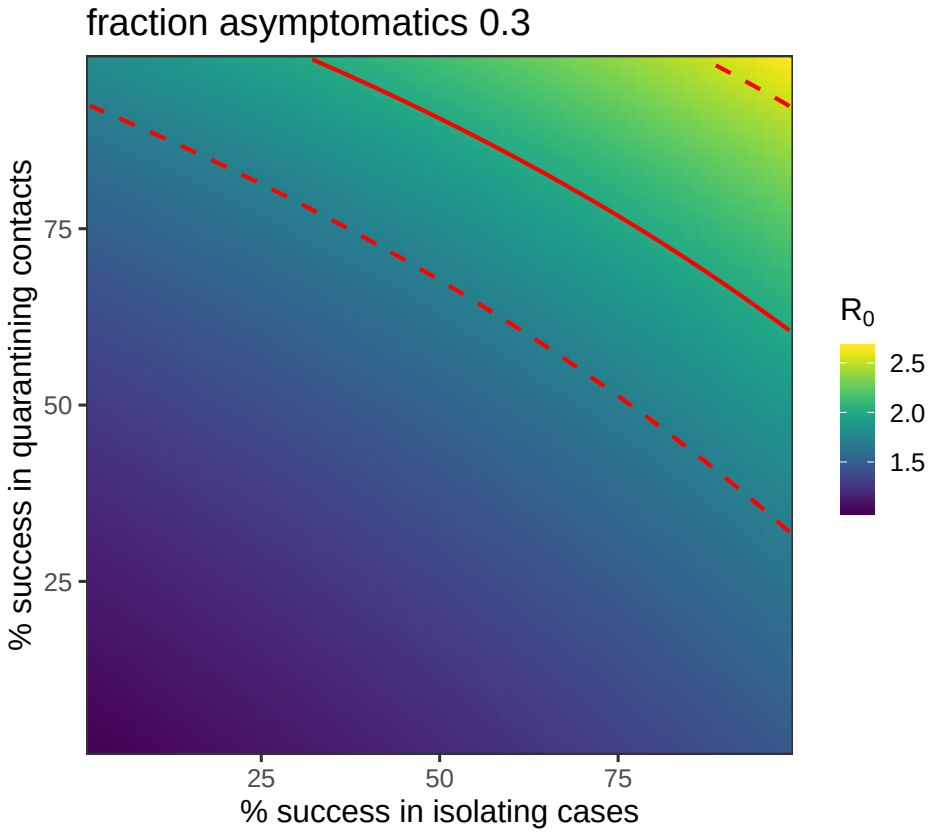
Supplementary Figure 9: Exactly as Figure 6 in the main paper (showing the strength of intervention needed to control the epidemic), but setting environmentally mediated contamination to zero.



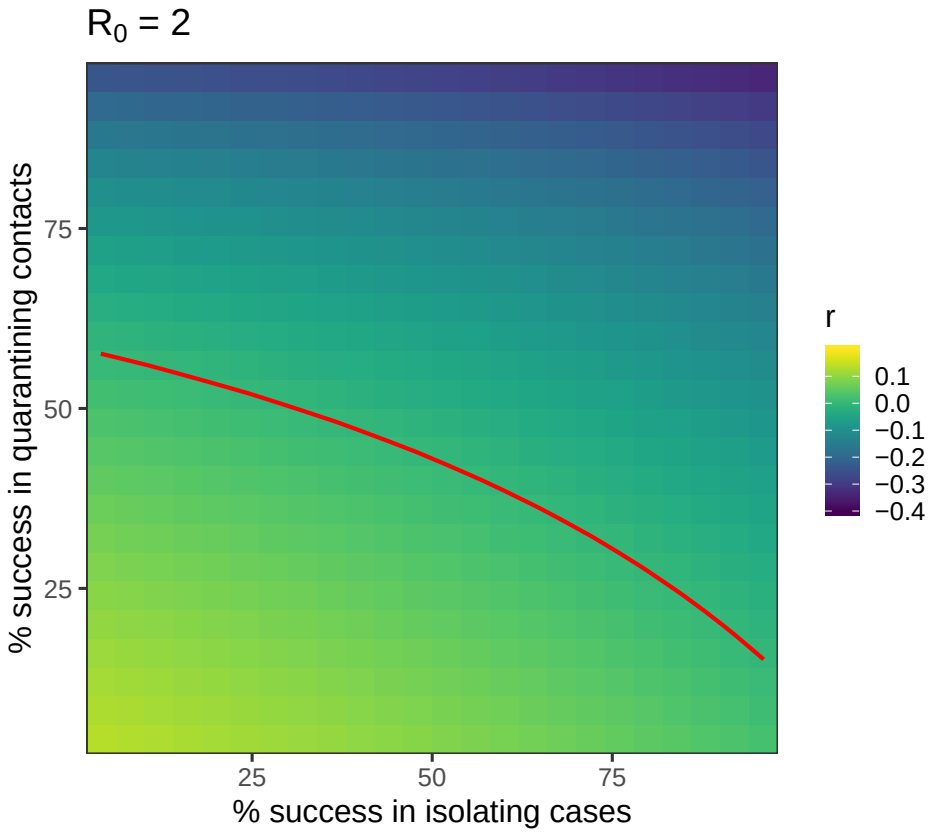
Supplementary Figure 10: As Supplementary Figure 9, but increasing the fraction of all transmission coming from asymptomatically infected individuals to 10%.



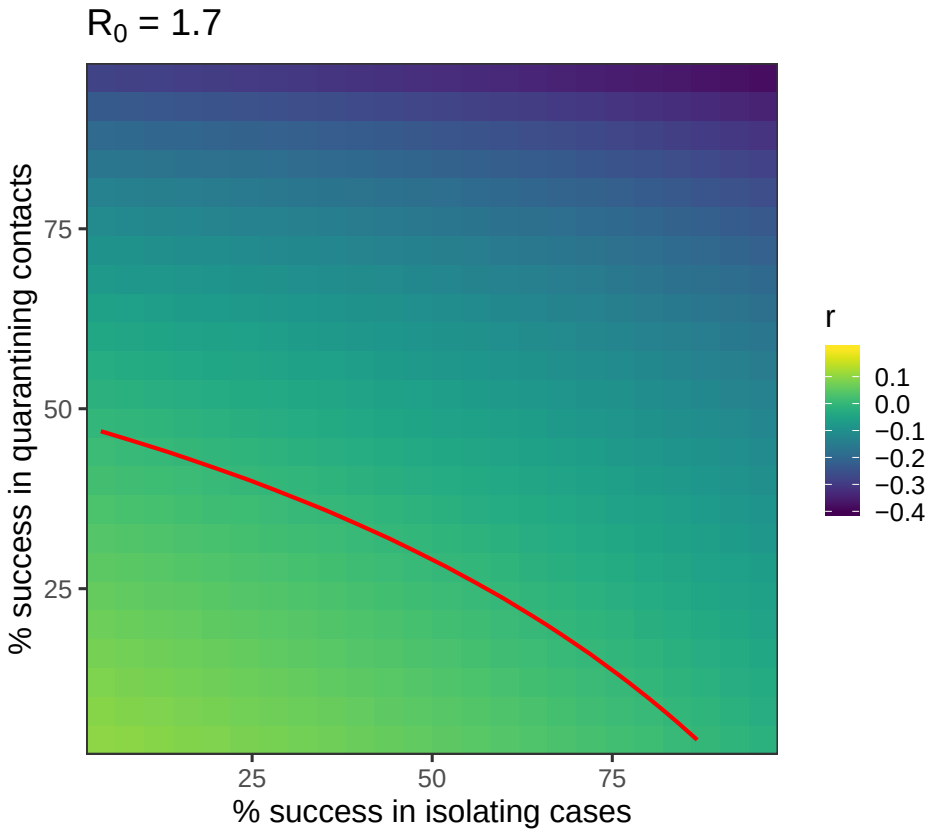
Supplementary Figure 11: As Supplementary Figure 9, but increasing the fraction of all transmission coming from asymptomatically infected individuals to 20%.



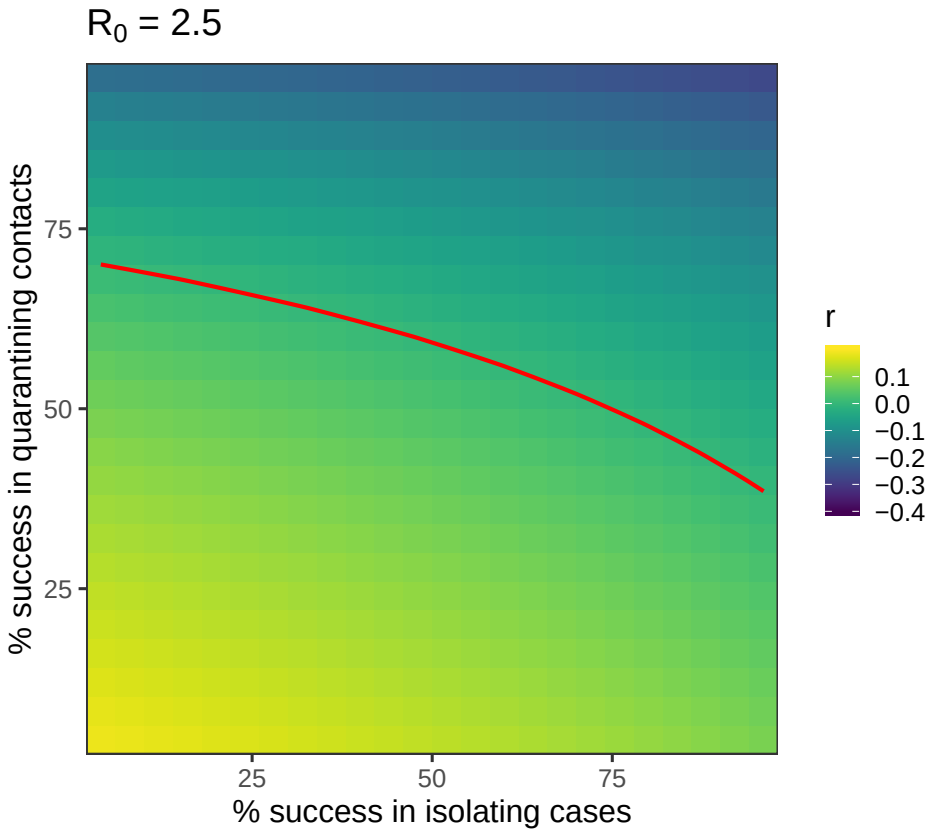
Supplementary Figure 12: As Supplementary Figure 9, but increasing the fraction of all transmission coming from asymptomatically infected individuals to 30%.



Supplementary Figure 13: Showing the exponential growth rate, r, resulting from application of case isolation and contact tracing with quarantining interventions. Epidemic control is achieved when r is zero, shown with a red line. With interventions weaker than those needed to achieve this, the epidemic still grows exponentially but at reduced rate compared to no intervention (with the bottom left point corresponding to no intervention). The figure assumes that asymptomatic and environmental transmissions are negligible.



Supplementary Figure 14: As Supplementary Figure 13 but fixing R_0 at the bottom end of the 95% CI, namely at 1.7.



Supplementary Figure 15: As Supplementary Figure 13 but fixing R_0 at the upper end of the 95% CI, namely at 2.5.