

# Quantifying the effects of population- and individual-based intervention strategies

Sang Woo Park<sup>1,\*</sup>, Kaiyuan Sun, Benjamin M. Bolker, Bryan T. Grenfell, Jonathan Dushoff, *[SWP: and maybe others if needed?]*

<sup>1</sup> Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

\*Corresponding author: swp2@princeton.edu

## 1 Introduction

*[SWP: Citations are missing in many places—I can fill them up easily later.] [SWP: Intros are always hard... will eventually re-write a lot of it but wanted to keep something there for now...]*

*[JD: Maybe suppress much of the early stuff about what can be measured or estimated directly. It's interesting that it would be easier to do  $\mathcal{R}$  directly in theory, but in practice  $\mathcal{R}$  is generally estimated indirectly and using similar assumptions. We can suppress this from the early Intro and get to it when we get to Cori.]*

The reproduction number  $\mathcal{R}$ —defined as the average number of secondary cases caused by a primary case—is a key characteristic of an emerging epidemic: It provides information about whether a disease can invade (i.e., when  $\mathcal{R} > 1$ ), the level of intervention required to prevent further spread, and the final size of an epidemic. When an epidemic is ongoing, transmission is affected by changes in population-level immunity and non-pharmaceutical interventions—these changes in transmission dynamics can be captured by  $\mathcal{R}(t)$ , often referred to as the *effective* or *time-dependent* reproduction number. Interpretation and estimation of  $\mathcal{R}(t)$  has been a key area of research during the current COVID-19 pandemic due to its policy implications.

*[SWP: This whole paragraph sounds repetitive... just trying to write something quickly to put all the ideas there.]* The main challenge in interpreting  $\mathcal{R}(t)$  can be attributed, in part, to its definition. Typically, it has been defined as the average number of secondary cases caused by a primary case in a partially susceptible population. While this definition sounds seemingly intuitive, it is mathematically imprecise. *[SWP: I wonder if it's better to leave out two sentences about imprecise-ness of the verbal definition? I sort of like to point it out though.]* Given infection kernel  $K(t, \tau)$  representing the rate at which an individual infected  $\tau$  time units ago generates secondary cases at time  $t$ , we can define time-dependent reproduction numbers in two different ways. First, the instantaneous reproduction number  $\mathcal{R}(t)$ —defined as the average number of secondary cases that an individual infected at time  $t$  will generate over the course of their infection if conditions at time  $t$  were to remain unchanged—corresponds to:

$$\mathcal{R}(t) = \int K(t, \sigma) d\sigma. \quad (1)$$

The instantaneous reproduction number measures conditions at time  $t$  and provides a real-time estimate of whether the disease will continue to spread if conditions stay the same; however, since conditions can continuously change throughout an epidemic,  $\mathcal{R}(t)$  cannot be measured directly (in theory), even if the entire infection process is known, and is difficult to interpret. On the other hand, we can also define the forward-looking case reproduction number  $\mathcal{R}_c(t)$ —defined as the average number of secondary cases that an individual infected at time  $t$  generated over the course of their infection:

$$\mathcal{R}_c(t) = \int K(t + \sigma, \sigma) d\sigma. \quad (2)$$

Unlike  $\mathcal{R}(t)$ ,  $\mathcal{R}_c(t)$  can be measured directly (in theory) by taking the average of secondary cases generated by each person and is easy to interpret; however,  $\mathcal{R}_c(t)$  depends on conditions after time  $t$  and therefore can only be estimated retrospectively.

[SWP: really just throwing out ideas here...] Estimating  $\mathcal{R}(t)$  adds another layer of challenge: It requires knowledge of true incidence of infection  $i(t)$  and the generation-interval distribution  $g(\tau)$ . Definition. Popularized by Cori et al, an estimator for  $\mathcal{R}(t)$  is given by:

$$\mathcal{R}(t) = \frac{i(t)}{\int_0^\infty g(\sigma) i(t - \sigma) d\sigma}. \quad (3)$$

Current practices generally assume that the generation-interval distribution  $g(\tau)$  does not vary over time.... This is good for capturing population-based intervention, but not individual-based intervention. For example, if a strict NPI (e.g., digital contact tracing) that prevents individuals from transmitting after symptom onset, then  $g$  would look really different... Neglecting such changes in  $g$  can introduce biases.

In this manuscript, we explore how population- and individual-based interventions can affect population- and individual-level measures of transmission. [SWP: Use  $r(t)$  as a complementary measure along with  $\mathcal{R}(t)$ !!!!]

## 2 Mathematical theory

### 2.1 Renewal equation framework

The renewal equation framework provides a flexible way of modeling the spread of infection and the impact of intervention on the spread. Let  $K(t, \tau)$  represent the infection kernel, defined as the rate at which secondary cases are generated at time  $t$  by an individual infected  $\tau$  time units ago (we will use  $s, t$  to denote calendar time and  $\sigma, \tau$  to denote time since infection throughout). The shape of this kernel can depend on internal factors (e.g., variation in infectiousness over the course of infection) or external factors (e.g., susceptible depletion, nonpharmaceutical interventions, changes in behavior). The integral of  $K(t, \tau)$ , representing the total infectiousness of an individual infected at time  $t$  given conditions at time  $t$ , corresponds to the instantaneous reproduction number:  $\mathcal{R}(t) = \int K(t, \sigma) d\sigma$ . The kernel, normalized by the total infectiousness, provides information about the time scale of

transmission; we refer to this quantity as the instantaneous generation-interval distribution:  
 $g(t, \tau) = K(t, \tau)/\mathcal{R}(t)$ . Then, incidence at time  $t$  caused by a cohort of individuals infected  
 $\tau$  time units ago can be written as the product of the instantaneous reproduction number  
 $\mathcal{R}(t)$ , the instantaneous generation-interval distribution  $g(t, \tau)$ , and incidence at time  $t - \tau$ ,  
 $i(t - \tau)$ :

$$i_{t-\tau}(t) = \mathcal{R}(t)g(t, \tau)i(t - \tau). \quad (4)$$

Integrating across time since infection  $\tau$  allows us to express the dynamics of the proportion  
susceptible  $S(t)$  and incidence  $i(t)$  in the absence of natural births or deaths as follows:

$$\frac{dS}{dt} = -i(t), \quad (5)$$

$$i(t) = \mathcal{R}(t) \int_0^\infty g(t, \sigma)i(t - \sigma) d\sigma. \quad (6)$$

For example, in the absence of intervention, the infection kernel  $K(t, \tau)$  can be written as a  
product of proportion susceptible  $S(t)$  and the intrinsic kernel  $K_0(\tau) = K(0, \tau)$ :

$$K(t, \tau) = S(t)K_0(\tau), \quad (7)$$

$$= \mathcal{R}_0 S(t)g_0(\tau), \quad (8)$$

where  $\mathcal{R}_0$  represents the basic reproduction number,  $g_0(\tau)$  represents the intrinsic generation-  
interval distribution. This model, also known as renewal equations, generalizes the dynamics  
of many compartmental models.

While renewal equations are often written in the form of Eq. 6, it is also useful to  
consider the forward-looking renewal equation because neither  $\mathcal{R}(t)$  nor  $g(t, \tau)$  are directly  
measurable (even under the assumption that all infection events can be observed); instead,  
both quantities rely on the assumption that the condition at time  $t$  remains unchanged.  
Consider the forward kernel  $F_t(\tau)$ , which represent the rate at which an individual infected  
at time  $t$  generates a secondary cases  $\tau$  time units after infection:

$$F_t(\tau) = K(t + \tau, \tau) = \mathcal{R}(t + \tau)g(t + \tau, \tau). \quad (9)$$

Unlike the infection kernel  $K(t, \tau)$ , the forward kernel  $F_t(\tau)$  is directly measurable, at least in  
theory. The integral of  $F(t, \tau)$ , representing the total infectiousness of an individual infected  
at time  $t$ , corresponds to the case reproduction number:  $\mathcal{R}_c(t) = \int F_t(\sigma) d\sigma$ ; this quantity  
can be theoretically measured by taking the average of number of infectees of individuals  
who were infected at time  $t$ . The kernel, normalized by the total infectiousness, correspond  
to the forward generation-interval distribution  $f_t(\tau) = F_t(\tau)/\mathcal{R}_c(t)$ ; this quantity can be  
theoretically measured by taking the distribution of time between actual infection events  
(i.e., realized generation intervals, given that the infector was infected at time  $t$ . At this  
point, it is straightforward to see that  $F_{t^*}(\tau) = K(t^*, \tau)$  (and therefore,  $\mathcal{R}_c(t^*) = \mathcal{R}(t^*)$   
and  $f_{t^*}(\tau) = g(t^*, \tau)$ ) if  $K(t, \tau)$  stays invariant after time  $t^*$  (i.e.,  $K(t, \tau) = K(t^*, \tau)$  for all  
 $t \geq t^*$ ); this hypothetical scenario, in which the condition at time  $t^*$  remains unchanged after  
time  $t^*$ , matches the core assumption behind the defition of instantaneous quantities  $\mathcal{R}(t)$

and  $g(t, \tau)$ . Then, incidence at time  $t$  caused by a cohort of individuals infected  $\tau$  time units ago can be written as:

$$i_{t-\tau}(t) = \mathcal{R}_c(t - \tau) f_{t-\tau}(\tau) i(t - \tau). \quad (10)$$

Integrating across time since infection  $\tau$  gives us the forward renewal equation:

$$\frac{dS}{dt} = -i(t), \quad (11)$$

$$i(t) = \int_0^\infty \mathcal{R}_c(t - \sigma) f_{t-\sigma}(\sigma) i(t - \sigma) d\sigma. \quad (12)$$

The forward renewal equation reveals an important insight: Even if the instantaneous generation-interval distribution  $g(t, \tau)$  remains invariant across time (i.e.,  $g(t, \tau) = g(0, \tau)$ ), changes in  $\mathcal{R}(t)$  causes the forward generation-interval to change (see Eq. 9). Therefore, renewal equation models that rely on the forward form (Eq. 12) but assume time-invariant  $f_t(\tau)$  do not reflect real epidemiological processes and must not be used.

## 2.2 Population- and individual-based interventions

To model the impact of intervention during an ongoing epidemic, we first distinguish population-based interventions, which reduces the transmission potential of all (or some fraction of) infected individuals by an equal amount regardless of their time of infection, from individual-based interventions, which target each individual separately and therefore depend on time of infection. Population-based interventions can be thought of as interventions that reduce the transmission rate and include social distancing, school closures, and vaccination. Individual-based interventions can be thought of as interventions that reduce the duration of infectious periods and include case isolation and behavioral changes (e.g., symptomatic individuals are more likely to self isolate earlier as awareness increases).

Let  $\mathcal{P}(t)$  represent a population-based intervention that modulates the infection kernel multiplicatively at calendar time  $t$  such that  $\mathcal{P}(t) = 1$  corresponds to intervention that has no effect and  $\mathcal{P}(t) < 1$  corresponds to intervention that reduces transmission. Then, the infection kernel under  $\mathcal{P}$  at calendar time  $t$  can be written as:

$$K(t, \tau) = \mathcal{R}_0 S(t) \mathcal{P}(t) g_0(\tau). \quad (13)$$

Likewise, the forward kernel of an individual infected at time  $t$  under  $\mathcal{P}$  can be written as:

$$F_t(\tau) = \mathcal{R}_0 S(t + \tau) \mathcal{P}(t + \tau) g_0(\tau). \quad (14)$$

Here, we see that population-based intervention has same impact on transmission dynamics as susceptible depletion.

For example, a social distancing measure that reduces transmission potential by a factor of  $1/\phi$  between time  $t_{\text{start}}$  and  $t_{\text{end}}$  can be modeled as:

$$\mathcal{P}(t) = \begin{cases} 1 & t < t_{\text{start}} \\ \phi & t_{\text{start}} \leq t < t_{\text{end}} \\ 1 & t_{\text{end}} \leq t \end{cases}. \quad (15)$$

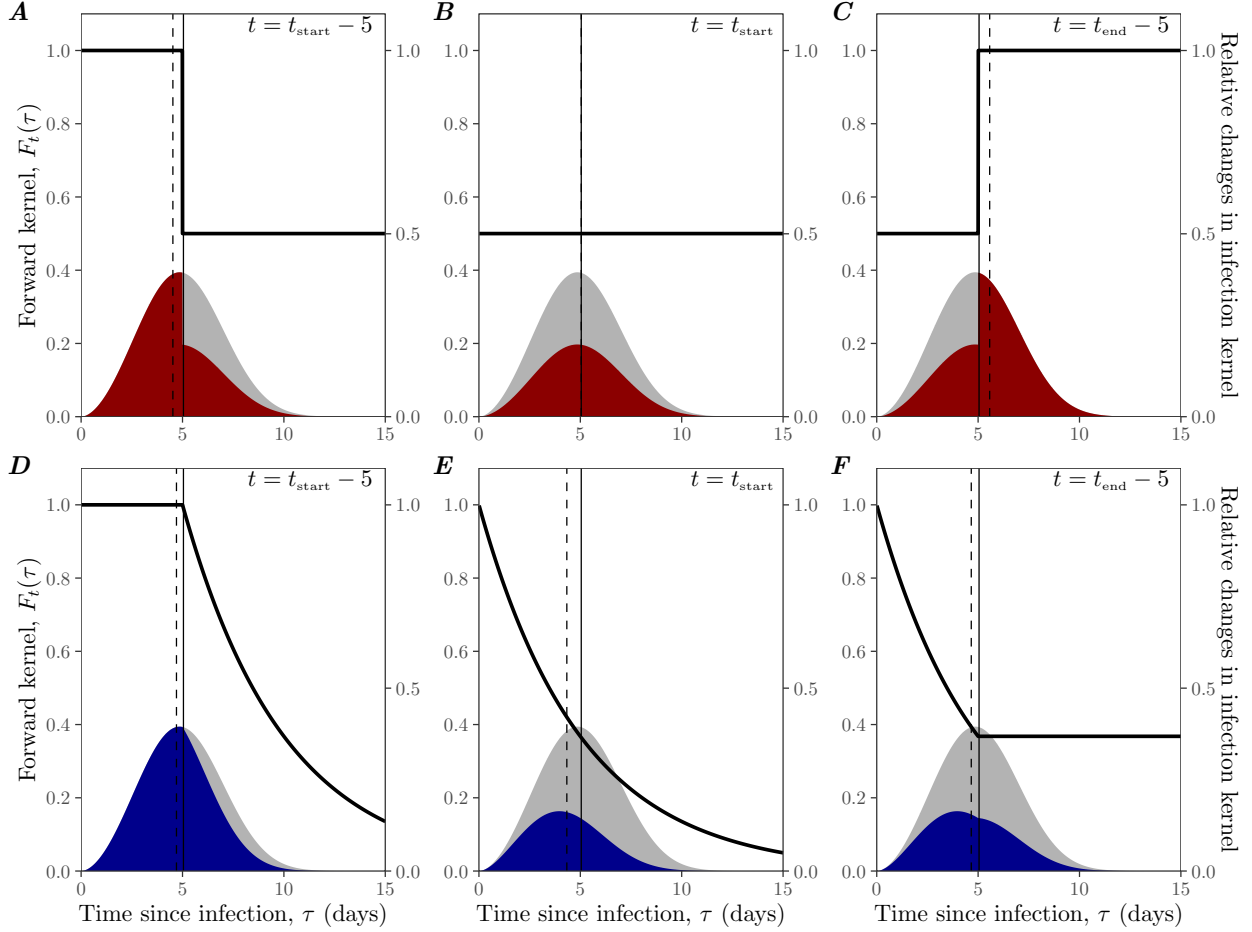


Figure 1: **The impact of population- and individual-based interventions on forward kernels.** The impact of constant strength (A–c) and speed (D–F) intervention on forward kernels of individuals infected at different time: 5 days before intervention onset (A, D), during intervention (B, E), and 5 days before intervention offset (C, F). Gray shaded curves represent the intrinsic kernel  $K_0(\tau)$ . Colored curves represent the forward kernel  $F_t(\tau)$ . Solid black lines represent relative changes in the kernel:  $F_t(\tau)/K_0(\tau)$ . Solid vertical lines the mean intrinsic generation interval. Dashed vertical lines the mean forward generation interval.

Fig. 1A–C illustrates the impact of such intervention on the forward kernel of an individual infected 5 days before  $t_{\text{start}}$ , at  $t_{\text{start}}$ , and 5 days before  $t_{\text{end}}$  (assuming  $S(t) \approx 1$ ). We present changes in the forward kernel due to its measurability. Population-based interventions can take effect immediately and sharply reduce transmission (Fig. 1A); likewise, lifting the intervention can, in theory, cause the forward kernel to return back to normal immediately (Fig. 1C). Even though the exact shape of the kernel depends on the time of infection and when the intervention was introduced relative to the infection time, the relative impact of intervention in reducing transmission at any given time (by a factor of  $1/\phi$ ) does not vary

across individuals (Fig. 1A–C). Such intervention also has direct effects on realized generation intervals: implementing (lifting) intervention decreases (increases) future transmission potential and therefore decreases (increases) the mean realized generation interval (Fig. 1A,C). If an individual is infected after  $t_{\text{start}}$  (and much earlier than  $t_{\text{end}}$ ), this intervention simply reduces the entire kernel by a constant amount and has no effect on realized generation intervals. This mechanism explains and further generalizes how susceptible depletion leads to contraction of generation intervals in a homogeneously mixing population.

On the other hand, individual-based intervention  $\mathcal{I}(t, \tau)$ , such as case isolation, targets each infected individual and therefore depends on calendar time  $t$  as well as time since infection  $\tau$ . In theory,  $\mathcal{I}(t, \tau)$  can capture almost all interventions; here, we use  $\mathcal{I}(t, \tau)$  to specifically model interventions that take the following form:

$$\mathcal{I}(t, \tau) = \exp \left( - \int_0^\tau h(t - \tau + \sigma, \sigma) d\sigma \right), \quad (16)$$

where  $h(t, \tau)$  represents the hazard rate at which an individual infected  $\tau$  time units ago will be isolated at calendar time  $t$ . Then,  $\mathcal{I}(t, \tau)$  can be interpreted as the probability that an individual infected  $\tau$  time units ago has not been isolated by calendar time  $t$  such that  $\mathcal{I}(t, \tau) < 1$  represents reduction in transmission. In this case, the infection kernel under  $\mathcal{I}$  at calendar time  $t$  is given by:

$$K(t, \tau) = \mathcal{R}_0 S(t) \mathcal{I}(t, \tau) g_0(\tau). \quad (17)$$

Likewise, the forward kernel of an individual infected at time  $t$  under  $\mathcal{I}$  can be written as:

$$F_t(\tau) = \mathcal{R}_0 S(t + \tau) \mathcal{I}(t + \tau, \tau) g_0(\tau). \quad (18)$$

This intervention is speed-like because its effectiveness depends on how fast we can isolate infected individuals.

For example, given hazard  $h(\tau)$  of being isolated, the probability that an individual infected at time  $t$  has not been isolated by  $\tau$  time units after infection given that the individual-based intervention takes place between time  $t_{\text{start}}$  and  $t_{\text{end}}$  depends on the amount of time the individual has been exposed to this intervention. If the individual was infected before  $t_{\text{start}}$ , they will not be isolated until after  $t_{\text{start}}$ . If the individual was infected after  $t_{\text{end}}$ , they will never be isolated by this intervention. Therefore, such probability can be modeled as:

$$\mathcal{I}(t + \tau, \tau) = \begin{cases} 1 & t < t_{\text{start}} - \tau \\ \exp \left( - \int_{\max(0, t_{\text{start}} - t)}^{\min(\tau, t_{\text{end}} - t)} h(\sigma) d\sigma \right) & t_{\text{start}} - \tau \leq t < t_{\text{end}} \\ 1 & t_{\text{end}} \leq t \end{cases} \quad (19)$$

a Fig. 1D–E illustrates the impact of a single individual-based intervention with constant hazard on the forward kernel of an individual infected 5 days before  $t_{\text{start}}$ , at  $t_{\text{start}}$ , and 5 days before  $t_{\text{end}}$  (assuming  $S(t) \approx 1$ ). Unlike the population-based intervention, individual-based intervention does not take effect immediately; that is, the rate at which an infected

individual generates a secondary case at time  $t_{\text{start}}$  (e.g.,  $F_t(t_{\text{start}} - t)$  in Fig. 1D; more generally,  $K(t_{\text{start}}, \tau)$ ) remains unaffected by the intervention because it takes time to identify and isolate infected individuals. On the other hand, when the intervention is lifted at  $t_{\text{end}}$ , the value of the kernel at calendar time  $t_{\text{end}}$  ( $F_t(t_{\text{end}} - t)$  in Fig. 1F; more generally,  $K(t_{\text{end}}, \tau)$ ) remains unchanged because some fraction of infected individuals have already been isolated. In general, individual-based interventions shorten realized generation intervals because they prevent late transmission (Fig. 1D–E).

Finally, given (possibly multiple) population- and individual-based interventions,  $P(t)$  and  $I(t, \tau)$ , the infection kernel can be written as:

$$K(t, \tau) = \mathcal{R}_0 S(t) \mathcal{P}(t) \mathcal{I}(t, \tau) g_0(\tau). \quad (20)$$

The corresponding forward kernel can be written as:

$$F_t(\tau) = \mathcal{R}_0 S(t + \tau) \mathcal{P}(t + \tau) \mathcal{I}(t + \tau, \tau) g_0(\tau). \quad (21)$$

Therefore, the dynamics of  $S(t)$  and  $i(t)$  can be now written as:

$$\begin{aligned} \frac{dS}{dt} &= -i(t), \\ i(t) &= \mathcal{R}_0 S(t) \mathcal{P}(t) \int_0^\infty \mathcal{I}(t, \sigma) g_0(\sigma) i(t - \sigma) d\sigma. \end{aligned} \quad (22)$$

## 2.3 Quantifying changes in time-dependent reproduction number

The impact of intervention is often measured by the instantaneous reproduction number  $\mathcal{R}(t)$ :

$$\mathcal{R}(t) = \mathcal{R}_0 S(t) \mathcal{P}(t) \int_0^\infty \mathcal{I}(t, \sigma) g_0(\sigma) d\sigma. \quad (23)$$

Since  $\mathcal{R}(t)$  measures conditions at time  $t$ , we expect to observe changes in  $\mathcal{R}(t)$  as soon as intervention is implemented. Therefore, it provides a real-time measure for whether the disease will continue to spread or not.

Estimating  $\mathcal{R}(t)$  depends on the instantaneous generation-interval distribution  $g(t, \tau)$ , rather than the intrinsic generation-interval distribution  $g_0(\tau)$ :

$$g(t, \tau) = \frac{K(t, \tau)}{\mathcal{R}(t)} = \frac{\mathcal{I}(t, \tau) g_0(\tau)}{\int_0^\infty \mathcal{I}(t, \sigma) g_0(\sigma) d\sigma}, \quad (24)$$

which can vary across time when individual-based interventions are present. By rearranging Eq. 6, we obtain the following estimator for  $\mathcal{R}(t)$ :

$$\mathcal{R}(t) = \frac{i(t)}{\int_0^\infty g(t, \sigma) i(t - \sigma) d\sigma}. \quad (25)$$

While this estimator is similar in form to previously proposed estimators, the main difference is that it allows for the underlying generation-interval distribution to vary to account for individual-based intervention. In particular, Cori et al have popularized the estimation of  $\mathcal{R}(t)$  via the R package EpiEstim, but it also assumes that the underlying generation-interval distribution does not change over time; such method accurately captures changes in  $\mathcal{R}(t)$  under population-based intervention, but not under individual-based intervention.

In general, when both population- and individual-based interventions are present during an ongoing epidemic, the classical estimator that assumes a time-invariant generation-interval distribution measures a slightly different quantity. Given incidence between time 0 and  $t - \epsilon$  for  $\epsilon > 0$ , we can calculate the proportional reduction  $p(t)$  in true incidence  $i(t)$  at time  $t$  compared to a counterfactual incidence value that we would observe at time  $t$  if all intervention measures were completely removed at time  $t$  (including release of isolated individuals):

$$p(t) = \frac{\mathcal{R}_0 S(t) \mathcal{P}(t) \int_0^\infty \mathcal{I}(t, \sigma) g_0(\sigma) i(t - \sigma) d\sigma}{\mathcal{R}_0 S(t) \int_0^\infty g_0(\sigma) i(t - \sigma) d\sigma} \quad (26)$$

$$(27)$$

By substituting true incidence  $i(t)$  in the numerator and rearranging, we obtain the classical expression for the estimator for  $\mathcal{R}(t)$ —which we refer to as  $\mathcal{R}_{\text{prop}}(t)$ —that depends on the proportion of the susceptible population and proportional reduction in incidence, rather than reduction in infectiousness:

$$\mathcal{R}_{\text{prop}}(t) = \mathcal{R}_0 S(t) p(t) = \frac{i(t)}{\int_0^\infty g(\sigma) i(t - \sigma) d\sigma}. \quad (28)$$

Despite its wide use, this estimator is difficult to interpret in general and does not necessarily serve as a threshold value.

## 2.4 Quantifying changes in time-dependent growth rate

The instantaneous reproduction number  $\mathcal{R}(t)$  measures whether the epidemic will *eventually* grow or decline if current conditions remain unchanged; counterintuitively, it does not tell us whether the epidemic is growing or not at a given moment. Interpreting  $\mathcal{R}(t)$  estimates is further impeded by the fact that conditions change during an epidemic. Instead, if we want to quantify whether incidence is increasing or decreasing, we can simply calculate the epidemic growth rate. Often growth rates are used to measure the initial exponential growth rate; here we generalize the idea by taking the per-capita growth rate:

$$r(t) = \frac{1}{i(t)} \frac{di(t)}{dt}. \quad (29)$$

Then, by definition, incidence grows when  $r(t) > 0$  (and vice versa). Although this measure seems tautological, it is the best available measure for capturing whether incidence is growing or not at a given moment.



## 2.5 Quantifying changes in the effective generation-interval distribution

In order to accurately estimate the instantaneous reproduction number, we have to be able to first quantify how the instantaneous generation-interval distribution  $g(t, \tau)$  changes across time  $t$ . The instantaneous distribution measures the infectiousness of an individual infected  $\tau$  time units ago at time  $t$  and therefore is different from the distribution of realized generation intervals (i.e., time between actual infection events). For example, if we were to take all realized generation intervals that end at time  $t$  and form a distribution, we obtain what is known as the backward generation-interval distribution,  $b_t(\tau)$ :

$$b_t(\tau) = \frac{i(t - \tau)K(t, \tau)}{\int_0^\infty i(t - \sigma)K(t, \sigma) d\sigma}, \quad (30)$$

$$= \frac{i(t - \tau)g(t, \tau)}{\int_0^\infty i(t - \sigma)g(t, \sigma) d\sigma}, \quad (31)$$

which is, in fact, a convolution of the instantaneous generation-interval distribution with previous incidence of infection.

Some studies have suggested substituting the forward distribution  $f_t(\tau)$ , including the forward serial-interval distribution, instead to estimate  $\mathcal{R}(t)$ :

$$\mathcal{R}_{\text{forward}}(t) = \frac{i(t)}{\int_0^\infty i(t - \sigma)f_{t-\sigma}(\sigma) d\sigma}. \quad (32)$$

However, this choice is also suboptimal. While both the forward generation-interval distribution  $f_{t-\tau}(\tau)$  and the instantaneous generation-interval distribution  $g(t, \tau)$  depend on the rate at which secondary cases at time  $t$  are generated by a primary case infected at time  $t - \tau$ ,  $K(t, \tau)$ , they are normalized by different quantities:

$$f_{t-\tau}(\tau) = \frac{K(t, \tau)}{\int_0^\infty K(t - \tau + \sigma, \sigma) d\sigma} \neq \frac{K(t, \tau)}{\int_0^\infty K(t, \sigma) d\sigma} = g(t, \tau). \quad (33)$$

The forward distribution  $f_{t-\tau}(\tau)$  is normalized by the average number of secondary cases caused by a primary case infected at time  $t - \tau$ . On the other hand, the instantaneous distribution  $g(t, \tau)$  is normalized by the sum of average number of secondary cases that an individual infected at time  $t - x$  will generate at time  $t$  (integrated across  $x$ ).

Therefore, even if we can observe all realized generation intervals throughout an epidemic, estimating the instantaneous distribution  $g(t, \tau)$  is not trivial. Even within a homogeneously mixing population, in which the disease is allowed to spread unchecked (and therefore  $g(t, \tau) = g_0(\tau)$  for all  $t$ ), aggregating all realized generation intervals underestimates the mean intrinsic generation interval due to susceptible depletion. Either a deconvolution or a novel statistical tool is needed to accurately estimate the instantaneous distribution  $g(t, \tau)$ . To our knowledge, no studies have accomplished this yet.

### 3 Example: Semi-mechanistic SIR model

In order to understand how population- and individual-based interventions affect disease spread, we use a semi-mechanistic SIR model to generate synthetic data and compare estimates of  $\mathcal{R}(t)$ :

$$\frac{dS}{dt} = -\beta(t)SI, \quad (34)$$

$$\frac{dI}{dt} = \beta(t)SI - \gamma(t)I, \quad (35)$$

$$\frac{dR}{dt} = \gamma(t)I, \quad (36)$$

where  $S$ ,  $I$ ,  $R$  represent proportion of individuals that are susceptible, infected, and removed (either due to recovery or isolation measures);  $\beta(t)$  represent time-varying transmission rate; and  $\gamma(t)$  represent time-varying removal rate. This model is semi-mechanistic because it allows  $\beta$  and  $\gamma$  to change over time but their changes are not necessarily driven by any particular mechanism. In this case, the instantaneous kernel can be written as:

$$K(t, \tau) = \beta(t)S(t) \exp \left( - \int_{t-\tau}^t \gamma(s) ds \right). \quad (37)$$

We see that changes in transmission rate  $\beta(t)$  and removal rate  $\gamma(t)$  correspond to population- and individual-based intervention. In particular, abrupt changes in  $\gamma(t)$  does not affect the incidence immediately due to delays in isolating individuals. Finally, the instantaneous reproduction number is given by:

$$\mathcal{R}(t) = \beta(t)S(t) \int_0^\infty \exp \left( - \int_{t-\tau}^t \gamma(s) ds \right) d\tau. \quad (38)$$

When  $\gamma(t) = \gamma(0)$ , we obtain a familiar form:  $\mathcal{R}(t) = \mathcal{R}_0 S(t)$ , where  $\mathcal{R}_0 = \beta(0)/\gamma(0)$ .

Here, we compare six different scenarios, in which intervention is introduced and is later partially/completely lifted (Fig. 2). In the first three scenarios, we consider smooth introduction and lifting of population-based intervention, modeled by changes in  $\beta$  while using fixed  $\gamma$  (Fig. 2A–C). In the last three scenarios, we consider sudden introduction and lifting of individual-based intervention, modeled by changes in  $\gamma$  while using fixed  $\beta$  (Fig. 2D–F). We refer to each pair of scenarios as *equivalent* population- and individual-based intervention scenarios because they give identical incidence curves, meaning that changes in  $\beta$  and  $\gamma$  are jointly unidentifiable (see Methods and Materials).

Epidemiological dynamics under equivalent population-based intervention are presented in Fig. 3. In this case, the instantaneous reproduction number  $\mathcal{R}(t)$  matches the proportional reproduction number  $\mathcal{R}_{\text{prop}}(t)$  (Fig. 3D–F). Likewise, the estimated  $\mathcal{R}(t)$  using **EpiEstim** matches the true  $\mathcal{R}(t)$ ; the initial overestimation of  $\mathcal{R}(t)$  is caused by left-censoring (Fig. 3D–F).

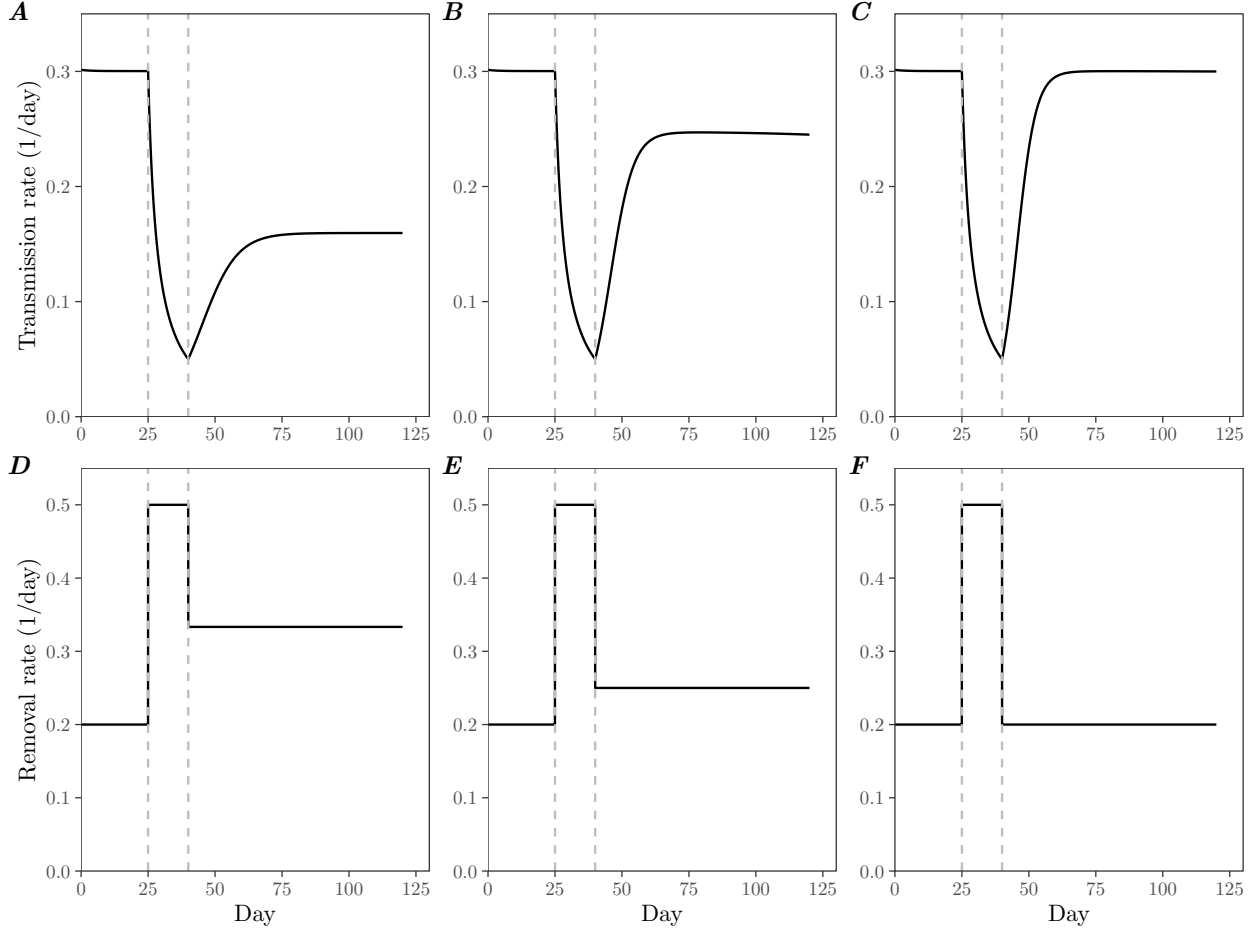


Figure 2: **Assumed changes in transmission rate and removal rate.** The semi-mechanistic SIR model is simulated under six different scenarios. (A–C) Equivalent population-based intervention scenarios. Removal rate  $\gamma$  is fixed to  $1/5 \text{ days}^{-1}$  throughout. (D–F) Equivalent individual-based intervention scenarios. Transmission rate  $\beta$  is fixed to  $3/10 \text{ days}^{-1}$  throughout. Each vertical pair of scenarios, referred to as equivalent population- and individual-based intervention scenarios, generates identical incidence curves.

## 4 Discussion

Speed is cool.

## 5 Methods

Here, we consider a simple scenario in which a flu-like pathogen with  $\mathcal{R}_0 = 1.5$  invades an immunologically naive population. In the beginning, the disease spreads without any intervention. On day 25, an intense case isolation measure is implemented. On day 40, the

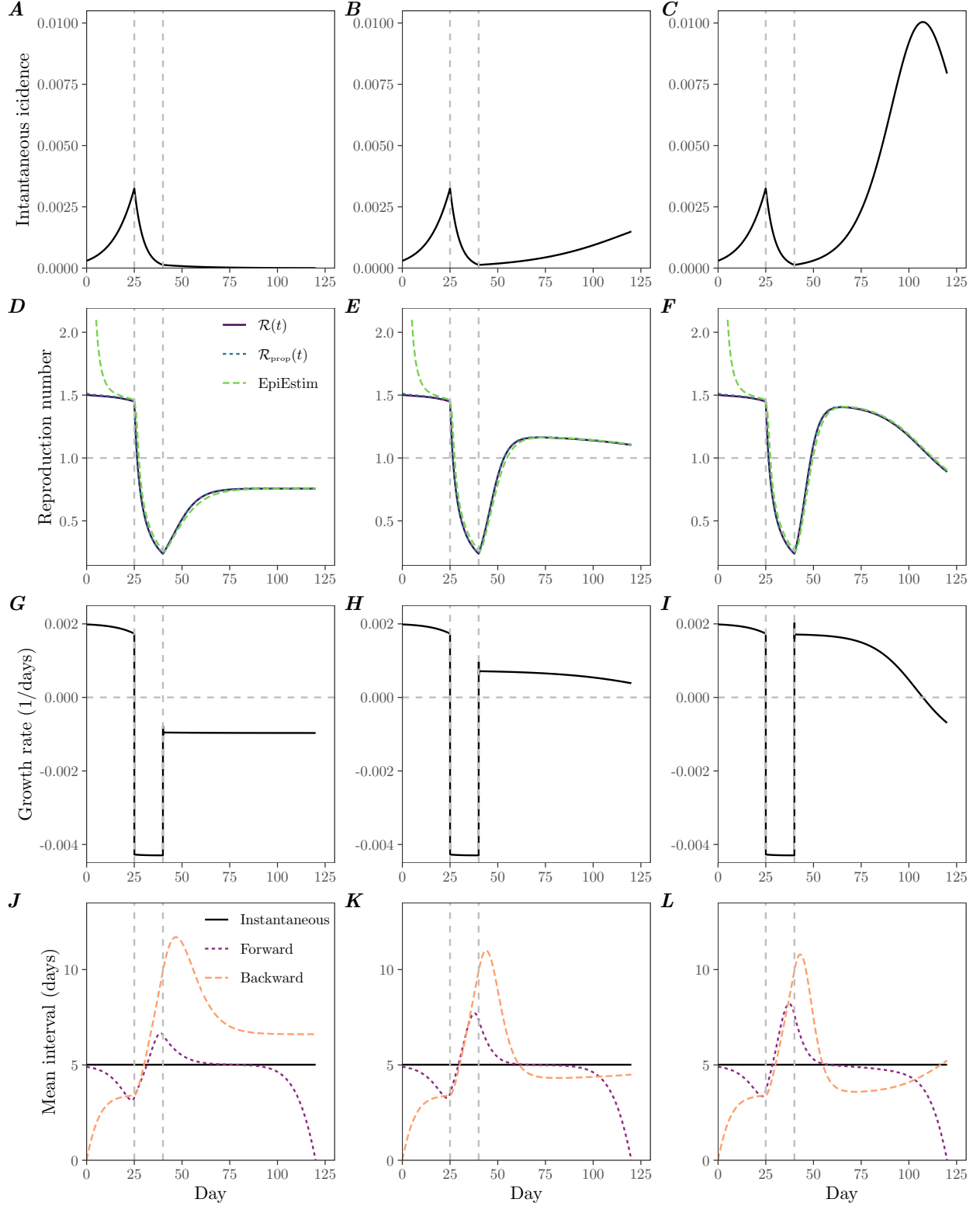


Figure 3: **Epidemiological dynamics of a semi-mechanistic SIR model under equivalent population-based intervention.** (A–C) Changes in instantaneous incidence over time. (D–F) Changes in instantaneous reproduction number  $\mathcal{R}(t)$ , proportional reproduction number  $\mathcal{R}_{\text{prop}}(t)$ , and estimated  $\mathcal{R}(t)$  using EpiEstim. (G–I) Changes in instantaneous growth rate. (J–L) Changes in mean instantaneous, forward, and backward generation intervals.

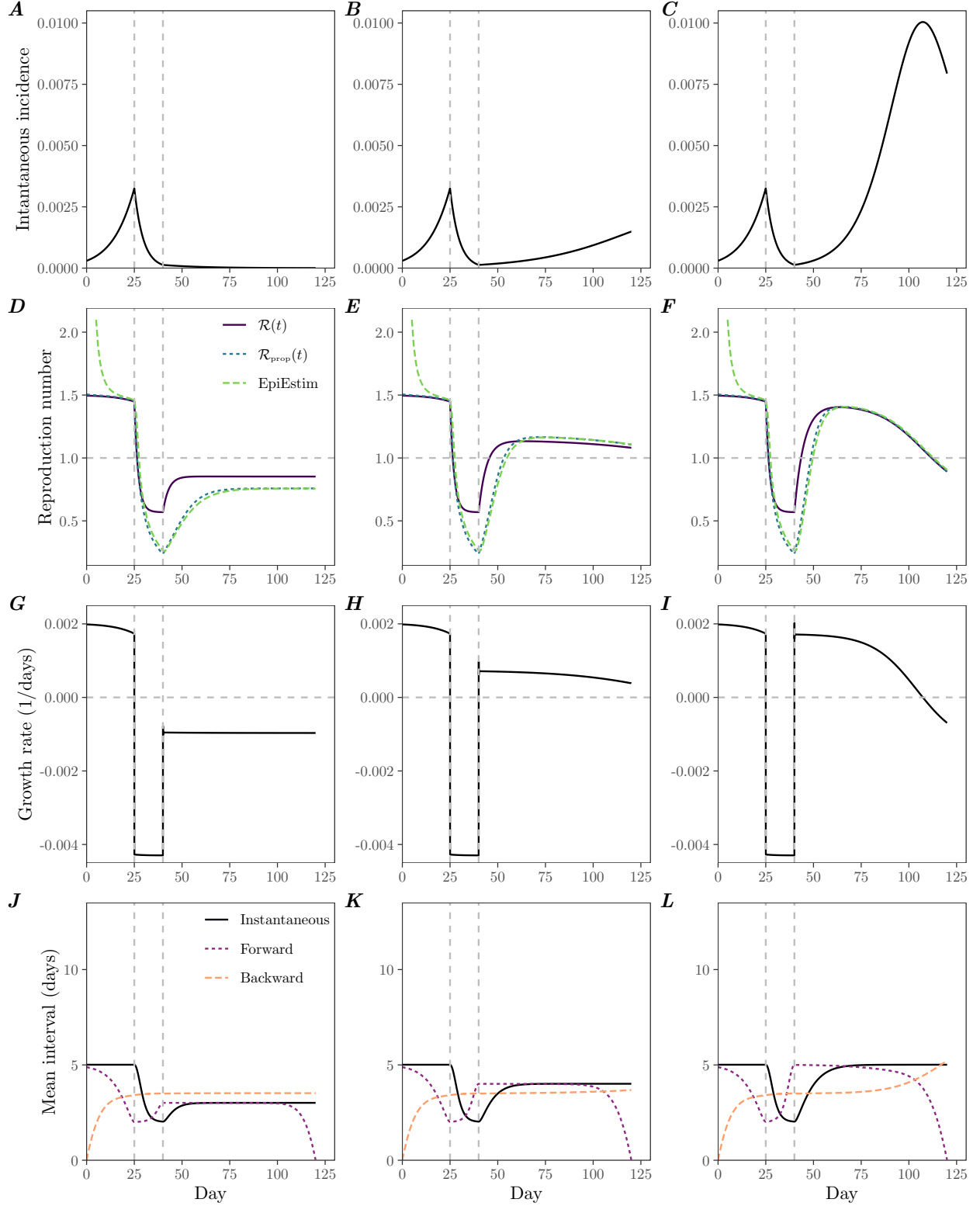


Figure 4: **Epidemiological dynamics of a semi-mechanistic SIR model under equivalent individual-based intervention.** (A–C) Changes in instantaneous incidence over time. (D–F) Changes in instantaneous reproduction number  $\mathcal{R}(t)$ , proportional reproduction number  $\mathcal{R}_{\text{prop}}(t)$ , and estimated  $\mathcal{R}(t)$  using EpiEstim. (G–I) Changes in mean instantaneous, forward, and backward generation intervals.

254 intervention is completely/partially lifted. This is modeled as follows:

$$\beta(t) = 3/10 \text{ days}^{-1}, \gamma(t) = \begin{cases} 1/5 \text{ days}^{-1} & t < 25 \\ 1/2 \text{ days}^{-1} & 25 \leq t < 40, \\ \gamma_{\text{late}} & 40 \leq t \end{cases} \quad (39)$$

255 where we vary  $\gamma_{\text{late}}$  between  $1/3 \text{ days}^{-1}$  (partial lifting) and  $1/5 \text{ days}^{-1}$  (complete lifting).  
 256 Simulations are run for 100 days based on the following initial conditions:  $S(0) = 1 - 10^{-3}$ ,  
 257  $I(0) = 10^{-3}$ , and  $R(0) = 0$ .

258 Since we assume that incidence is known exactly until time  $t^*$ , we can, in fact, estimate  
 259 the transmission rate  $\beta^*(t)$  (with fixed  $\gamma(t) = \gamma(0)$ ) that gives identical incidence trajectory  
 260 until time  $t^*$ . This transmission rate is given by:

$$\beta^*(t) = \frac{\mathcal{R}_{\text{prop}}(t)\gamma(0)}{S(t)}. \quad (40)$$

261 More generally, given true incidence  $i(t)$  and true  $\mathcal{R}(t)$  until time  $t^*$ , modulated by some  
 262 combination of population- and individual-based intervention, we can find population inter-  
 263 vention  $\mathcal{P}^*(t)$  that gives identical incidence until time  $t^*$ :

$$\mathcal{P}^*(t) = \frac{\mathcal{R}_{\text{prop}}(t)}{S(t)}. \quad (41)$$

264 We refer to this intervention as the *equivalent* population-based intervention.