

Research



Cite this article: Favero M, Scalia Tomba G, Britton T. 2022 Modelling preventive measures and their effect on generation times in emerging epidemics. *J. R. Soc. Interface* **19**: 20220128.
<https://doi.org/10.1098/rsif.2022.0128>

Received: 16 February 2022
Accepted: 12 May 2022

Subject Category:

Life Sciences—Mathematics interface

Subject Areas:

biomathematics

Keywords:

epidemic modelling, preventive measures, estimation bias, generation time, reproduction number

Author for correspondence:Martina Favero
e-mail: martina.favero@math.su.se

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6011452>.

Modelling preventive measures and their effect on generation times in emerging epidemics

Martina Favero¹, Gianpaolo Scalia Tomba² and Tom Britton¹¹Department of Mathematics, Stockholm University, Stockholm, Sweden²Department of Mathematics, University of Rome Tor Vergata, Roma, Italy

MF, 0000-0002-7069-9721; GST, 0000-0001-6581-8661; TB, 0000-0002-9228-7357

We present a stochastic epidemic model to study the effect of various preventive measures, such as uniform reduction of contacts and transmission, vaccination, isolation, screening and contact tracing, on a disease outbreak in a homogeneously mixing community. The model is based on an infectivity process, which we define through stochastic contact and infectiousness processes, so that each individual has an independent infectivity profile. In particular, we monitor variations of the reproduction number and of the distribution of generation times. We show that some interventions, i.e. uniform reduction and vaccination, affect the former while leaving the latter unchanged, whereas other interventions, i.e. isolation, screening and contact tracing, affect both quantities. We provide a theoretical analysis of the variation of these quantities, and we show that, in practice, the variation of the generation time distribution can be significant and that it can cause biases in the estimation of reproduction numbers. The framework, because of its general nature, captures the properties of many infectious diseases, but particular emphasis is on COVID-19, for which numerical results are provided.

1. Introduction

While the reproduction number is usually seen as a dynamic quantity, changing over the course of an epidemic, the generation time distribution is often seen as a static object. For example, during the recent COVID-19 pandemic, major efforts have been made to continuously estimate reproduction numbers, while often outdated estimates of the generation time distribution have been employed in the estimation. The distribution of the generation time, which we define here as the time between the infection of a secondary case and the infection of the corresponding primary case, is not an intrinsic property of an infectious disease; on the contrary, it depends on the environment and on the behaviour of the individuals among whom the disease spreads. Consequently, the generation time distribution can be subject to variations, for example, recent studies hint that generation times of SARS-CoV-2 have significantly shortened during the course of the pandemic [1–4]. When doing inference, it is important to investigate the extent of these variations and take them into account, when substantial.

The significance of the generation time distribution stems from its relation to the initial exponential growth rate of the epidemic and the basic reproduction number, or equivalently the current growth/decline rate and the current reproduction number. In fact, by observing the incidence, the growth rate is often estimated and used together with the generation time distribution to derive the reproduction number by means of the Euler–Lotka equation (see §3.2 for more details). Therefore, using a current estimate of the growth rate together with a generation time distribution estimated before preventive measures

were in place could lead to a biased estimate of the reproduction number. Estimating the reproduction number R in an ongoing epidemic is crucial for efficient control of the epidemic. If for example $R = 1.25$, the overall number of contacts of infectious individuals must be reduced by at least 20% to stop the epidemic from increasing ($(1 - 0.2)R = 1$). Similarly, a community currently having a high level of restrictions and current reproduction number $R = 0.8$ may relax restrictions as long as the number of contacts by infectious individuals does not increase by more than 25% ($(1 + 0.25)R = 1$). Biases in estimates of R may hence lead to improper conclusions regarding control measures.

Several issues arise in connection with estimation of the generation time distribution. To begin with, in practice, it is usually not possible to observe all of the secondary cases caused by a cohort of primary cases, and therefore biases due to missing cases arise. Furthermore, moments of infection are rarely observed and surrogate time points are used, such as time of symptom onset. When the epidemic is exponentially growing, short generation times are overrepresented among the observed ones. The generation time may vary during the course of the epidemic because of significant changes in the number of susceptible individuals, for example near the peaks of incidence. When performing inference, it is important to take the above mentioned issues into account (e.g. [5–9]). Another problem, of a different nature, is that the generation time may vary because of changes in the behaviour of individuals, in particular, due to preventive measures introduced to reduce the spread of the disease.

The latter problem has received some attention, see for example [1,3,4] in connection with the COVID-19 pandemic, but, unlike the other mentioned problems, has not been extensively studied yet. It is not uncommon that studies, which focus on the important task of estimating other quantities, use an outdated estimate of the generation time distribution which does not represent the current situation, because of changes in the level of preventive measures among different time periods or among different locations. For example, many influential and highly cited studies, such as [10,11], use the estimate of the generation time distribution of SARS-CoV-2 from early 2020 in Shenzhen, China [12]. See also [2], where this problem is highlighted, and references therein. Other influential studies, e.g. [13], use instead estimates of the generation time distribution [14] of SARS-CoV-1 which despite being similar to SARS-CoV-2, resulted in milder preventive measures.

The choice of using outdated estimates of the generation time distribution to obtain much-needed estimates of other quantities, such as reproduction numbers, is understandable; however, it may lead to biases. That is why in this paper, we focus on illustrating, both theoretically and in a COVID-19 example, how various preventive measures change the generation time distribution in emerging epidemics and we draw attention to some scenarios in which it would be preferable to replace the initial estimate of the generation time distribution with an up-to-date estimate that takes into account the non-negligible effect of interventions. A related problem is the effect of asymptomatic individuals on generation times, which we briefly discuss in §4.2.

We focus on the initial phase of an epidemic, that is, when depletion of susceptible individuals is negligible. As the epidemic progresses, some of the potentially infectious contacts of infected individuals will be with already infected or

recovered (assumed immune) individuals and thus the effective reproduction number will decrease. The reduction is proportional to the fraction of individuals that have already been infected and hence only slowly changing in the initial phase of the epidemic. Determining the length of this phase depends on the specific outbreak, and thus goes beyond the scope of this paper, but can be done for example by observing that the infection rate in the beginning of the epidemic is close to constant up until a certain percentage of individuals have been infected, say 5%. The time it takes to reach this threshold identifies the initial phase. Our analysis of the generation time distribution can also be easily transferred to other phases of the epidemic in which the fraction of susceptible individuals does not change significantly. Continuing the previous example, also later in the epidemic the infection rate is close to constant over a period where at most 5% get infected. This period is usually shorter around the peak; for example, it might take a few months for the first 5% to get infected and later in the epidemic, around the peak, a few weeks for 5% more to get infected. Note that considering later phases of the epidemics would allow one to take into account naturally acquired immunity, but not waning of immunity if waning is observed over a longer period than the periods mentioned above.

We present, in §2, a general stochastic epidemic model for the spread of infectious diseases with a structure that is particularly indicated to analyse the impact of preventive measures on the distribution of generation times and reproduction numbers. The key feature of the model that facilitates this analysis is a random infectivity profile with a random time at which the contact activity of an infectious individual is reduced. We define this random time to be equal to the time of symptoms onset, when no preventive measures are in place, and equal to the time of detection (that can occur by various means, including symptoms onset), when interventions are in place. We assume that for each individual the rate of symptoms onset is proportional to infectiousness; see §4.1 for a discussion on this assumption and further details.

In particular, we consider the following interventions: homogeneous reduction of contacts and transmission, vaccination, isolation, screening and contact tracing. Clearly, all these preventive measures aim at reducing the reproduction number; however, not all of them have an effect on generation times. The analysis in this paper shows that homogeneously reducing infectivity, by reducing contacts or transmission, as well as vaccination do not change the generation time distribution; see §§5.1 and 5.2. On the contrary, reducing the contact activity of, or isolating, symptomatic individuals does have a significant impact on generation times; see §5.3. For example, in the COVID-19 scenario illustrated in §6, the natural intuition that this preventive measure shortens generation times is confirmed. Furthermore, our analysis shows that other interventions affecting the generation time distribution are those aiming at expediting the time at which an infectious individual is isolated, or reduces their contact activity, such as screening and contact tracing. In particular, in the COVID-19 scenario with isolation of detected individuals, if the population is screened uniformly, that is, each individual is randomly tested at a certain constant rate, generation times are shortened, and if contact tracing is put in place, the shortening becomes substantial. Using a pragmatic approach, aimed at providing a general analysis

Table 1. Notation summary.

infectivity process and infectivity function	$\lambda = \{\lambda(t)\}_{t \geq 0}, \beta$
infectiousness process	$X = \{X(t)\}_{t \geq 0}$
contact process	$C = \{C(t)\}_{t \geq 0}$
contact rate before and after symptom onset/ detection	C_1, C_2
time of contact activity reduction and its conditional rate	τ, α_τ
time of symptoms onset, its conditional rate and corresponding parameter	T_S, α_S, a_S
probability of asymptomatic infection	p^{asy}
time of detection and its conditional rate	T_D, α_D
time of screening and its rate	$T_{scre}, \alpha_{scre}, \sigma$
time of contact tracing and its conditional rate	T_{CT}, α_{CT}
probability of successful contact tracing	p
other quantities related to contact tracing	$\alpha_{CT1}, f, \alpha_{CT2},$ a_{CT2}, d
reduction fractions of contacts and transmission	ρ_G, ρ_x
reduction fractions due to vaccination and to isolation of symptomatic and detected individuals	$\rho_V,$ ρ_S, ρ_D
fraction of vaccinated individuals	v
relative susceptibility (vaccination response)	A, a
relative infectivity (vaccination response)	B, b

of the impact of contact tracing at the population level, we make an approximation of the contact tracing mechanism and we provide a rigorous derivation of the rate at which individuals are contact traced. See §5.4 and the electronic supplementary material for a detailed discussion on screening and contact tracing.

In §5.5, we summarize the results in a general formula for the generation time distribution that takes into account the cumulative effect of all interventions. The model, the analysis and the general formula for the generation time distribution (§§2–5) are valid under general conditions and thus can be used to study various infectious diseases by defining a suitable infectivity profile and parameters. In §6, we tune the model to fit a COVID-19 scenario to illustrate the theoretical general results in a realistic example, and, perhaps more importantly, to investigate the extent of the impact of interventions on the generation time distribution and reproduction number in a framework that allows for further extensions.

2. The stochastic model

The model we present can be interpreted as a special case of the very general epidemic models in e.g. [8,15], tailored to consider preventive measures and to study their impact. A summary of the notation, including stochastic processes, random variables, parameters and functions, is given in table 1. We adopt the classical assumption that the

population is homogeneously mixing. This is a natural assumption, despite it being a simplification of reality, as it allows an analysis which is valid for the numerous wide-spread models that are based on the very same assumption. Future investigation could consist of removing this assumption and repeating the analysis for more complex models based on networks. In this model, we also assume that all individuals are equally susceptible at the start of the epidemic, whereafter they might be infected and later removed. On the contrary, infectious individuals are not equally infectious, they have infectivity profiles which are independent and identically distributed, as in [8,15]. That is, the infectivity profile of an infectious individual is the realization of a stochastic process λ , the infectivity process, which is the product of a contact process and an infectiousness process, described in the following. This allows modelling some inhomogeneity in the population; in fact, although identically distributed, individuals' contact behaviours and infectiousness can vary. It is known [15] that, when the initial phase of an epidemic is considered, and thus depletion of susceptible individuals is negligible, this epidemic model corresponds to a Crump–Mode–Jagers branching process [16]. This allows using results from the theory of branching processes to analyse the epidemic model.

2.1. Infectiousness process

The infectiousness profile of an infectious individual is an independent realization of the continuous-time, $[0, 1]$ -valued stochastic process $X = \{X(t)\}_{t \geq 0}$, the infectiousness process. Given X , if a contact between the infectious individual with infectivity profile X and a susceptible individual occurs at time t (since the infection of the infectious individual), then infection of the susceptible individual happens with probability $X(t)$. In the following, we mention two possible choices for the infectiousness process.

Example 2.1 (Infectiousness profiles with a deterministic shape). Let X_1 be a random variable with values in $[0, 1]$, X_2 a random variable with values in $\mathbb{R}_{>0}$, and let $h: \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ be a deterministic function. Then a possible construction of the infectiousness process is $X(t) = X_1 h(tX_2^{-1})$. This is based on the assumption that the infectiousness profile has the same shape for all individuals, given by the function h . Furthermore, since not all individuals are equally infectious, each individual is associated with a pair of random variables distributed as (X_1, X_2) , which determine the strength and the location of the peak of infectiousness.

Example 2.2 (Infectiousness profiles in the SIR model). In the well-known SIR stochastic model, the infectiousness process corresponds to $X(t) = x \mathbf{1}_{[0, I]}(t)$, where x is the probability of infection and I is the (random) length of the infectious period.

2.2. Contact process

The contact activity of an infectious individual is a realization of the continuous-time, $\mathbb{R}_{\geq 0}$ -valued, stochastic process $C = \{C(t)\}_{t \geq 0}$, the contact process. An individual's contact activity may depend on their infectiousness profile, but is

assumed to be independent of other individuals' contact activities. Given C , an infectious individual with contact activity C has contacts with other individuals at the time points of an inhomogeneous Poisson process with intensity $C(t)$, time being measured since the infection of the infectious individual.

We assume that the contact process has the following form:

$$C(t) = \begin{cases} C_1, & \text{if } t \leq \tau \\ C_2, & \text{if } t > \tau \end{cases} = C_1 1_{t \leq \tau} + C_2 1_{t > \tau}, \quad (2.1)$$

where τ is a random time at which the initial base contact activity C_1 is reduced to C_2 . The $\mathbb{R}_{\geq 0}$ -valued random variables C_1 and C_2 , with $C_2 \leq C_1$, are possibly dependent on each other, but are assumed to be independent of anything else. We assume that the random time τ has conditional rate $\alpha_\tau(t)$, at time t , given the contact activities, C_1 , C_2 , and given the infectiousness processes until time t , that is, given \mathcal{X}_t which corresponds to the natural filtration associated with the infectiousness process. This means that the conditional distribution of τ is of the form

$$G_\tau(t) = \mathbb{P}(\tau > t \mid \mathcal{X}_t, C_1, C_2) = \exp\left(-\int_0^t \alpha_\tau(u) \, du\right). \quad (2.2)$$

Having defined a contact process of the form (2.1) allows enough flexibility to describe realistic scenarios we are interested in analysing, while enabling a solid mathematical formulation. In fact, as shown in the next sections, different definitions of α_τ , C_1 , C_2 allow modelling contacts in several scenarios, with and without preventive measures. In particular, τ can be defined to model for example the time of symptoms onset or, more generally, the time of detection by symptoms onset or by other preventive measures, e.g. screening and contact tracing. More details on τ are provided in §§4 and 5.4, where modelling of symptoms onset and detection are, respectively, addressed.

2.3. Infectivity process

Infectivity profiles of infectious individuals are independent realizations of the continuous-time, $\mathbb{R}_{\geq 0}$ -valued stochastic process $\lambda = \{\lambda(t)\}_{t \geq 0}$, with

$$\lambda(t) = C(t)X(t).$$

The infectivity process λ combines the effects of the infectiousness process and the contact process. In a susceptible population, given λ , an infectious individual with infectivity profile λ infects other individuals at the time points of an inhomogeneous Poisson process with intensity $\lambda(t)$, time being measured since the infection of the infectious individual. This is equivalent, by properties of Poisson processes, to the description in the previous subsections: the infectious individual has contacts at the time points of an inhomogeneous Poisson process with intensity $C(t)$, and a contact at time t results in an infection with probability $X(t)$. Note that the process that counts the infectious contacts, i.e. the counting process $N = \{N(t)\}_{t \geq 0}$, with random intensity λ , is a Cox process, also known as doubly stochastic Poisson process (e.g. [17]). We have described the role of the infectivity process in a completely susceptible population; instead, if some individuals are not susceptible, contacts with those cannot result in infection.

However, in the initial phase of an epidemic, which is what this paper focuses on, depletion of susceptible individuals is negligible and thus the above description of the infectivity process is valid.

3. Some quantities of interest

3.1. Reproduction numbers

In this framework, the infectivity function, i.e. the average rate at which a typical infected individual infects others in a fully susceptible population, is simply the expectation of the infectivity process, that is,

$$\beta(t) = \mathbb{E}[\lambda(t)] = \mathbb{E}[C_1 X(t) G_\tau(t) + C_2 X(t)(1 - G_\tau(t))], \quad (3.1)$$

where $G_\tau(t)$ is defined in (2.2). Therefore, the average number of individuals infected by a typical infectious individual, the reproduction number, is

$$R = \int_0^\infty \beta(t) \, dt = R^{(1)} + R^{(2)}, \quad (3.2)$$

where

$$R^{(1)} = \mathbb{E}\left[\int_0^\infty C_1 X(t) G_\tau(t) \, dt\right]$$

is the average number of individuals infected by a typical infectious individual before reducing their base contact activity (e.g. before symptoms onset, or before detection), and

$$R^{(2)} = \mathbb{E}\left[\int_0^\infty C_2 X(t)(1 - G_\tau(t)) \, dt\right]$$

is the average number of individuals infected by a typical infectious individual after changing their contact activity.

The above formulas for R , $R^{(1)}$ and $R^{(2)}$ give a theoretical means of comparison of the effect of various preventive measures; in particular, considering $R^{(1)}$ and $R^{(2)}$ separately can be useful to better understand strengths and limits of a certain intervention. For example, by simply isolating infectious individuals after they show symptoms, $R^{(2)}$ can be significantly reduced; however, if no other preventive measures are in place, $R^{(1)}$ remains unchanged and thus R cannot be reduced below $R^{(1)}$. However, often in practice, the reproduction number cannot be estimated directly by formulae of the type above and estimates rely instead on another approach, which is described in the following subsection.

3.2. Generation time distribution

A valuable tool to estimate the reproduction number is the Euler–Lotka equation which relates the reproduction number R to the, usually more easily observable, Malthusian parameter, the exponential growth rate r (e.g. [18]). As shown in [5], in the initial phase of an epidemic, the incidence quickly approaches exponential growth, with rate r , and thus the Euler–Lotka equation, arising from population dynamics, applies:

$$1 = R \int_0^\infty e^{-rt} g(t) \, dt.$$

This equation links the reproduction number and the

growth rate through the function g , the generation time distribution, which, in the framework of this paper, as in [8], is expressed as

$$g(t) = \frac{\mathbb{E}[\lambda(t)]}{\mathbb{E}[\int_0^\infty \lambda(u) du]} = \frac{\beta(t)}{R}. \quad (3.3)$$

It is because of its appearance in the Euler–Lotka equation that the generation time distribution has a crucial role in inference. In fact, knowing g allows deriving R from r or vice versa. As pointed out in [8], the intuitive definition of the generation time as the time between a primary and a secondary case has several mathematical counterparts, and it is thus important to recall that g in the Euler–Lotka equation corresponds to the following definition. Consider a cohort of infectious individuals (referred to as primary cases) and all of their secondary cases. Choose one of the secondary cases uniformly at random and measure the time between the infection of the secondary case and the infection of the corresponding primary case. The probability density function of this random time is equal to g in (3.3); see [8] for more details. Note that this sampling procedure is size-biased, that is, a primary case associated with more secondary cases is more likely to be chosen compared to one associated with fewer secondary cases. For this reason, choosing first a primary case uniformly at random and then one of its secondary cases (also uniformly at random) would yield a different generation time distribution, $\mathbb{E}[\lambda(t)/\int_0^\infty \lambda(u) du]$, which is not of interest here and should not be confused with (3.3).

This paper focuses on the study of variations of the generation time distribution due to preventive measures. Let us explain why these variations are worth being studied. Imagine an epidemic evolving without preventive measures, with basic reproduction number R_0 , generation time distribution g_0 and growth rate r_0 . When interventions are introduced, R_0 , g_0 , r_0 are replaced by R_E , g_E , r_E . Each of the triplets above satisfies the Euler–Lotka equation. When doing inference before interventions, r_0 is observed and R_0 is derived using g_0 . The same should be done when various interventions are put in place, that is, r_E and g_E give R_E . However, if we observed r_E and used g_0 instead of g_E , the resulting reproduction number would be biased. The aim of this paper is to study how the generation time distribution varies, from g_0 to g_E , when preventive measures are introduced, and how that affects the estimate of the reproduction number.

4. Modelling symptoms and the neutral scenario

The stochastic model described in §2 can capture several scenarios, including the neutral scenario in which no preventive measures are in place. In this case, the random time τ , at which the contact activity of an individual is reduced, is assumed to be equal to the time of symptoms onset T_S . In fact, a natural reduction of the contact activity can occur at symptoms onset, even without preventive measures, for example, even if isolation of symptomatic people is not required, having flu-like symptoms naturally reduces contact activity.

Furthermore, by defining $\tau = T_S$, we can study all those interventions that do not affect the time at which the contact reduction occurs; see §§5.1–5.3. Whereas, for those

interventions that aim at expediting the time at which the contact reduction occurs, e.g. by implementing other means of detection besides symptoms onset, τ needs to be defined differently, i.e. as the time of detection; see §5.4.

This section is dedicated to $\tau = T_S$. In particular, §4.1 is dedicated to mathematical definition of T_S , the underlying biological assumptions, and the resulting formulae, which will be crucial in the following sections, while §4.2 contains a comment on the effect of asymptomatic cases.

4.1. Random time of symptoms onset

Let T_S be the time between infection and symptoms onset of an infectious individual. Consider the model of §2 with $\tau = T_S$. We assume that T_S has conditional rate, given \mathcal{X}_t ,

$$\alpha_S(t) = a_S X(t), \quad (4.1)$$

where $a_S \in \mathbb{R}_{>0}$. In this way, the distribution of the time to symptoms onset of an individual depends on their infectiousness process and is independent of their contact process. Furthermore, the above assumption implies that, at any given time, the higher the infectiousness, the higher the rate of symptoms onset. Therefore, it is more likely for a symptomatic infectious individual to show symptoms in proximity to the peak of their infectiousness, in line with what is observed for COVID-19 (e.g. [19,20]), and with the viral load of patients with pandemic H1N1 2009 virus infection peaking on the day of onset of symptoms [21]. Furthermore, note that in this model, an infectious individual is not necessarily symptomatic, even if highly infectious. We call an infectious individual who never shows symptoms asymptomatic (to not be confused with an individual who has not shown symptoms yet, called pre-symptomatic); this corresponds to $T_S = \infty$. The probability that an individual with infectiousness process X is asymptomatic is

$$\mathbb{P}(T_S = \infty | X) = e^{-\int_0^\infty a_S X(u) du}.$$

This means that the higher the total infectiousness, $\int_0^\infty X(u) du$, the higher the probability of the individual being symptomatic. More generally, the probability that a typical individual is asymptomatic is

$$p^{\text{asy}} = \mathbb{P}(T_S = \infty) = \mathbb{E} \left[e^{-\int_0^\infty a_S X(u) du} \right].$$

Making different biological assumptions on symptoms onset would correspond to a different mathematical definition of T_S . For example, T_S could be defined as the time at which the infectiousness process exceeds a certain threshold. From now on (4.1) is assumed. This is not only a realistic assumption but also mathematically convenient, yielding explicit formulae for $R^{(1)}$ and $R^{(2)}$, the average number of individuals infected by a typical infectious individual before and after symptoms onset, respectively. See the electronic supplementary material for more details.

4.2. Biases due to asymptomatic individuals

Usually, the generation time distribution is estimated from data related to symptomatic cases. That is, instead of the real generation time distribution g , the generation time distribution of symptomatic cases, g^{sy} , is estimated, and in turn used to estimate the reproduction number, leading to biases (e.g. [4,22]). In this subsection, we briefly analyse how these two distributions are related.

Let $\beta^{\text{asy}}(t) = \mathbb{E}[\lambda(t) \mid T_S = \infty]$ and $\beta^{\text{sy}}(t) = \mathbb{E}[\lambda(t) \mid T_S < \infty]$ be the average infectivity rates of asymptomatic and symptomatic individuals, respectively. Then the average infectivity rate is

$$\beta(t) = p^{\text{asy}}\beta^{\text{asy}}(t) + (1 - p^{\text{asy}})\beta^{\text{sy}}(t).$$

Let $R^{\text{asy}} = \int_0^\infty \beta^{\text{asy}}(t) dt$ and $R^{\text{sy}} = \int_0^\infty \beta^{\text{sy}}(t) dt$ be the average numbers of individuals infected by a typical asymptomatic infectious individual and by a typical symptomatic individual, respectively. It is straightforward to calculate that

$$R = p^{\text{asy}}R^{\text{asy}} + (1 - p^{\text{asy}})R^{\text{sy}}.$$

The formula above confirms the obvious intuition that the higher the fraction p^{asy} of asymptomatic individuals, the higher the impact of asymptomatic individuals and thus, when data on asymptomatic individuals are unavailable, estimates of R might be biased, unless a correction is made. Furthermore, $g^{\text{sy}}(t) = (1/R^{\text{sy}})\beta^{\text{sy}}(t)$ and the generation time distribution of asymptomatic individuals is $g^{\text{asy}}(t) = (1/R^{\text{asy}})\beta^{\text{asy}}(t)$, thus

$$g(t) = q^{\text{asy}}g^{\text{asy}}(t) + (1 - q^{\text{asy}})g^{\text{sy}}(t),$$

where

$$q^{\text{asy}} = \frac{p^{\text{asy}}R^{\text{asy}}}{p^{\text{asy}}R^{\text{asy}} + (1 - p^{\text{asy}})R^{\text{sy}}}$$

indicates how much asymptomatic transmission affects the overall generation time distribution.

5. Interventions

By varying the contact process C , the infectiousness process X , or the random time τ , several types of intervention, and their effect on generation times and reproduction number, can be studied. In particular, the interventions analysed in this paper are grouped in the following categories:

- homogeneous reduction of contact level (e.g. physical distancing, lockdown),
- homogeneous reduction of transmission probability (e.g. face masks),
- vaccination,
- isolation of symptomatic individuals,
- screening, and
- contact tracing.

Note that the latter two interventions aim at expediting the time at which an individual is detected and at lowering their contact activity after detection, while the remaining ones focus on lowering the infectivity process, without affecting the detection time. In the following subsections, we analyse the different types of preventive measures, considering one at a time and comparing it with the no-interventions neutral scenario. In §5.5, the cumulative effect of all interventions is considered and a general formula is provided.

5.1. Homogeneous reduction of contact activity or transmission probability

This subsection is dedicated to the analysis of two types of preventive measures that lead to the reduction of the infectivity process by a multiplicative factor.

The first class of preventive measures that we consider consists of those measures, such as physical distancing or lockdown, that reduce the contact process by a factor ρ_C , which is a $[0, 1]$ -valued random variable. This means that each individual reduces their contact activity by a factor which is an independent realization of ρ_C .

The second class instead consists of those measures, such as introducing face masks, that reduce the infectiousness process by a factor ρ_X , which is a $[0, 1]$ -valued random variable. This means that the infectiousness profile of each individual is reduced by a factor which is an independent realization of ρ_X .

These measures are homogeneous in the population, that is, we are assuming that each individual is recommended/required to follow the same measures independently of their situation. The random variables ρ_C and ρ_X are thus independent of the infectivity process. While the measures are homogeneous, the individual response is variable, and thus we use random variables, instead of deterministic constants, to represent the variability in the personal adherence.

Both classes of measures, albeit in different ways, have the same effect on the infectivity process, which is reduced by a multiplicative factor, i.e. $\lambda_E(t) = \rho_C\rho_X\lambda(t)$, which yields $\beta_E(t) = \mathbb{E}[\rho_C\rho_X]\beta(t)$ and $R_E = \mathbb{E}[\rho_C\rho_X]R_0$. Therefore, the preventive measures considered in this subsection affect the reproduction number, while leaving the generation time distribution unchanged.

5.2. Vaccination

In this subsection, we complement the underlying epidemic model with a vaccination model as in e.g. [23,24]. We assume that a fraction v of the population receives a vaccine, before the epidemic starts. While this is a simplification of reality, it allows an analysis of the generation time distribution in a population that is partly vaccinated and a comparison with the scenario without interventions. Assume that each vaccinated individual has a random response to the vaccine, determining the reduction in susceptibility and (if infected) infectivity. The response is described by the $[0, 1]$ -valued random variables A and B , the relative susceptibility and the relative infectivity respectively. This means that a vaccinated individual with response A, B has a probability of getting infected reduced by A compared to the probability of getting infected without vaccine, and if ever infected, their infectivity is reduced by a factor B . Note that we are assuming that the vaccine may reduce the infectiousness profile by a multiplicative factor B , without changing its shape. See the electronic supplementary material for a discussion on this assumption.

As in the previous subsection, the infectivity function, as well as the reproduction number, is reduced by a multiplicative factor, $\mathbb{E}[\rho_V]$, and therefore the generation time distribution is unchanged. For more details, see the electronic supplementary material.

Immunity, or partial immunity, from disease exposure has the same type of effect as vaccination, that is, it reduces the reproduction number without changing the generation time distribution. In fact, as explained in the Introduction, in this paper, we are considering a period of time in which the fraction of immune individuals does not change significantly; thus the reasoning around the effect of immune individuals is identical to the one around vaccinated individuals.

Finally, we remark that the effect of immunity waning in vaccinated individuals is not considered here. This goes beyond the scope of the model, since we only focus on dynamics over shorter time periods, i.e. the phases mentioned in the Introduction, rather than on the long-term evolution which usually needs to be considered when analysing the effect of waning.

5.3. Isolating symptomatic individuals

As mentioned in §4, in a scenario without preventive measures, the initial contact activity C_1 of an individual is naturally reduced to C_2 at the time of symptoms onset, T_S . If symptomatic individuals are recommended or required to isolate, then the contact activity after symptoms onset, C_2 , is further reduced by a factor $\rho_S \in [0, 1]$, ideally close to 0. This preventive measure can be easily included in the model and its effect on the generation time distribution and the reproduction number can be studied by analysing (3.1) and (3.2).

The basic reproduction number, in the no-interventions scenario is $R_0 = R_0^{(1)} + R_0^{(2)}$. Now assume that symptomatic individuals are required/recommended to isolate, while no other preventive measures are in place. It is straightforward to see that the reproduction number becomes $R_E = R_0^{(1)} + \rho_S R_0^{(2)}$. Therefore, by simply isolating symptomatic individuals, the reproduction can be lowered to a minimum of $R_0^{(1)}$, which corresponds to a scenario in which symptomatic individuals are completely isolated and do not have any contact with others. In fact, isolating symptomatic individuals has no impact on the amount of pre-symptomatic or asymptomatic transmission, which defines $R_0^{(1)}$. The generation time distribution, by (3.1) and (3.2), becomes

$$g_E(t) = \frac{1}{R_0^{(1)} + \rho_S R_0^{(2)}} \mathbb{E}[C_1 X(t) G_{T_S}(t) + \rho_S C_2 X(t) (1 - G_{T_S}(t))].$$

Therefore, isolating symptomatic individuals not only affects the reproduction number, but also, unlike the previously analysed interventions, changes the generation time distribution.

If other detection measures are in place, the preventive measure discussed in this subsection could be improved by isolating detected individuals in addition to symptomatic individuals. See the next subsections for more details.

5.4. Screening and contact tracing

In the previous subsections, the time at which an individual reduces their contact activity, τ , was assumed to coincide with the time of symptoms onset. While isolating symptomatic individuals or generally reducing the infectivity process homogeneously in the whole population does not affect τ , other preventive measures instead aim at reducing τ .

Interventions such as screening and contact tracing act by expediting the time at which an infectious individual is discovered to be infectious. This scenario can be modelled by letting $\tau = T_D$, the time of detection, and

$$T_D = \min\{T_S, T_{\text{scre}}, T_{\text{CT}}\},$$

where T_{scre} is the time of screening, T_{CT} is the time of detection by contact tracing. In this framework, we assume that, given the infectiousness and the contact process up to time t , i.e. given \mathcal{X}_t and C_t , the conditional rate of detection at time t since infection is

$$\alpha_D(t) = \alpha_S(t) + \alpha_{\text{scre}}(t) + \alpha_{\text{CT}}(t), \quad (5.1)$$

where α_S , the conditional rate of symptoms onset, is defined in (4.1), α_{scre} , the screening rate, is defined in (5.2) below, and α_{CT} , the conditional rate of detection by contact tracing, is defined in (5.3) below.

We assume screening is performed as follows. Random tests are carried out so that the entire population, including infectious and non-infectious individuals, is screened uniformly. That is, independently of their infectiousness and contact process, each individual is tested at a constant rate σ . At time t , given $X(t)$, the rate of detection by screening is thus

$$\alpha_{\text{scre}}(t) = \sigma 1_{X(t)>0}, \quad (5.2)$$

as an individual can only be detected when infectious. This means that, if we consider a small time period of length ϵ , any individual, infectious or not, has approximately a probability $\epsilon\sigma$ of being tested during that period, thus roughly a fraction σ of the entire population is tested each day, if time is measured in days. Uniform screening is not often used in practice, and thus not representative of most real-world scenarios. The next section illustrates that the effect of uniform screening is moderate, unless σ is large, which in practice requires a substantial screening effort. This motivates why uniform screening is not often implemented as a preventive measure. In order to obtain a bigger impact, it is more efficient to direct testing towards individuals that are more likely to be infectious, as in contact tracing programmes, rather than spreading it uniformly over the entire population. Being a more efficient strategy, this is also more representative of the real-world practice. For this reason, we now include contact tracing in the model.

Modelling contact tracing is notoriously challenging; see [25] for an extensive overview. Various modelling approaches are possible; see for example [26,27] for compartmental models, [3] for deterministic integral equations, [28] for an individual-based stochastic model; and important observations on data have been made (e.g. [29]). Nevertheless, simpler models that are wide-spread in practical applications rely on strong simplifying assumptions, while more accurate complex models are often intractable in practice. A complicating factor is that, because of contact tracing, infectious individuals are not independent of each other; for example, the infectivity profile of an infector might be truncated because one of their infectees develops symptoms quickly and triggers contact tracing leading to detection of the infector. In order to exactly model contact tracing, it is necessary to keep track of the status of each single individual and of relations between individuals by building additional mathematical structure in the model. An example is given in [28], where each individual in a stochastic SIRS model is associated with an id-number and with the id-number of their infector. This leads to complicated calculations and large simulations, even if the underlying epidemic model is rather simple. The heterogeneity of individuals, which is modelled in this paper by the stochastic infectivity process, leads to further difficulties. Therefore, instead of using a similar approach, aiming at exact expressions and keeping track of each single individual, we focus on approximating the effect of contact tracing to provide insight at population level.

The first approximation that we make is to assume that the infectivity profiles of different individuals are independent of each other, despite contact tracing. Each individual

can be contact traced either through one of their infectees, with conditional rate α_{CT1} , or through their infector, with conditional rate α_{CT2} . As a result of the approximation and of the assumptions described below, the rate at which an infectious individual is detected through contact tracing at a certain time t since their infection, given C_1 and the infectiousness profile up to time t , i.e. given \mathcal{X}_t , is

$$\begin{aligned} \alpha_{CT}(t) &= \alpha_{CT1}(t) + \alpha_{CT2}(t) \\ &= p \int_0^{t-d} C_1 X(u) f(t-d-u) du + pa_{CT2} 1_{X(t)>0}, \end{aligned} \quad (5.3)$$

where p is the probability that contact tracing occurs successfully, $pa_{CT2} 1_{X(t)>0}$ is the rate, approximated by a constant, at which an infectious individual is detected through their infector, f is the probability density function of the time between the moments an infectee is infected and detected and d is the (deterministic) contact tracing delay. It is implied that, if $t \leq d$, the integral in (5.3) is equal to zero and thus $\alpha_{CT1}(t) = 0$. We assume that contact tracing stops after one step, which is a reasonable approximation of reality. In fact, it is unlikely that an individual is contact traced through their infectee who in turn has been contact traced, because this would usually take longer than the infectious period. This latter assumption allows deriving an explicit expression for the function f , which would not be possible otherwise; see the electronic supplementary material. Formula (5.3) can be easily generalized to include a random contact tracing delay, as shown in the electronic supplementary material. Finally, assuming that the infector is detected at a constant rate is also an approximation of reality. In fact, this rate varies with time and depends on how long has passed since the infector was infected, while remaining independent of the infectivity profile of the infectious individual under consideration. A non-approximated derivation of this rate is challenging, as explained in the electronic supplementary material, and would require additional structure to be added to the model, which goes beyond the scope of this paper and is left for future work.

5.5. The cumulative impact of all interventions

Finally, to summarize the analysis of this section, we consider altogether the preventive measures that have been analysed so far and provide a general formula for the generation time distribution under the effect of all preventive measures. When all preventive measures are in place, and in particular, not only symptomatic individuals but also detected individuals reduce their contact activity by a fraction ρ_D , the infectivity function becomes

$$\begin{aligned} \beta_E(t) &= \mathbb{E}[\rho_V \rho_C \rho_X] \\ &\cdot \mathbb{E}[C_1 X(t) G_{T_D}(t) + \rho_D C_2 X(t) (1 - G_{T_D}(t))], \end{aligned} \quad (5.4)$$

with

$$\begin{aligned} G_{T_D}(t) &= \exp\left(-a_S \int_0^t X(u) du - (\sigma + pa_{CT2})(t \wedge I) \right. \\ &\quad \left. - pC_1 \int_0^{t-d} X(u) [1 - y(t-d-u)] du\right) \end{aligned} \quad (5.5)$$

and

$$y(t) = \mathbb{E}\left[\exp\left(-\int_0^t a_S X(u) du - \sigma(t \wedge I)\right)\right], \quad (5.6)$$

where \wedge indicates the minimum and I is the length of the infectious period, i.e. $I = \int_0^\infty 1_{X(u)>0} du$. From the expressions

above, it is clear that, while all interventions affect the reproduction number, $R_E = \int \beta_E(u) du$, only some interventions are found to have an impact on the generation time distribution, $g_E(t) = \beta_E(t)/R_E$, that is, isolation, screening and contact tracing. In §6, the expression above is used to illustrate variations of the generation time distribution in a realistic example.

To relate our results to a model used in numerous applications, we conclude this section with two examples concerning the well-known SIR model, which can be seen as a special case of our general model.

Example 5.1 (Basic SIR model). Continuing with the SIR model of example 2.2, we recall that the infectivity process is of the form $\lambda(t) = c\xi 1_{[0,I]}$, where ξ is the probability of infection, c is the contact rate and I is the length of the infectious period, which is exponentially distributed with parameter a_I . In this framework, it is straightforward to show that $g(t) = a_I \exp(-a_I t)$, that is, the generation time is exponentially distributed with parameter a_I (e.g. [9]).

Example 5.2 (SIR model with interventions). Consider a SIR model where the infectious period can be cut short by symptoms onset. That is, the length of the infectious period is $I' = \min\{I, T_S\}$. The time of symptoms onset has rate $\alpha_S(t) = a_S \xi 1_{[0,I]}$, thus $G_{T_S}(t) = \exp(-a_S \xi(t \wedge I))$. It is then easy to show that the generation time is exponentially distributed with parameter $a_S \xi + a_I$. Note that in example 5.1 neither c nor ξ influence the generation time distribution. On the contrary, in this example, a higher ξ lowers the mean generation time by increasing the rate of symptoms onset, $a_S \xi$.

Screening and contact tracing can be also considered in the SIR model and a formula for the generation time distribution can be easily obtained from formulae (5.4)–(5.6), by simply plugging in $X(t) = \xi 1_{[0,I]}$, using that I is exponential and analytically computing the integrals. Straightforward but lengthy calculations show that the generation time distribution in this example is a generalization of a truncated (positive) Gumbel distribution. See the electronic supplementary material for more details.

6. Illustration: COVID-19 outbreak

In this section, we tune the model to resemble a COVID-19 outbreak and illustrate the impact of preventive measures in such a framework. While we use available evidence to define the parameters of the model and thus provide a realistic illustration of COVID-19 scenarios, we do not directly use data in our study. The results presented here are not to be considered as proper estimates of generation times and reproduction numbers for the COVID-19 pandemic, but rather as a means to illustrate the extent of variation caused by interventions.

The infectiousness process is assumed to be of the form described in example 2.1 with the function $h(t)$ being a Gamma density, with shape 2.5 and rate 0.5, shifted by 2, in line with the analysis in [20,30], and X_1 and X_2 are uniformly distributed in [0.1, 1.9] and [0.5, 1.5], respectively. We recall that each infectious individual has an infectiousness profile which corresponds to a realization of the random infectivity profile, thus, in this case, to a realization of the variables X_1 and X_2 . In figure 1, some infectiousness profiles,

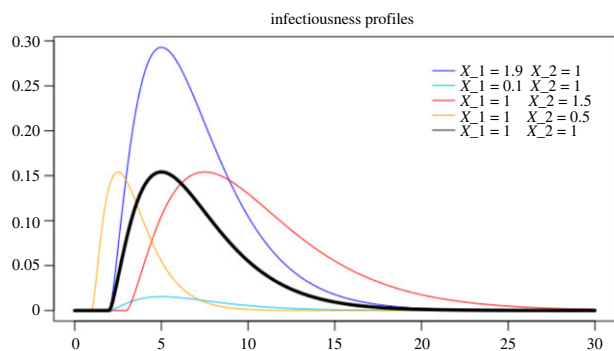


Figure 1. In black, the deterministic curve that shapes the infectiousness process. In colours, various infectiousness profiles, corresponding to the most extreme cases: in blue and light blue, the highest and lowest peak of infectiousness, respectively; in red and orange, the latest and earliest peaks of infectiousness, respectively.

corresponding to different realizations of X_1 and X_2 are plotted. In particular are plotted the underlying profile, corresponding to $X_1=1$, $X_2=1$, and the four profiles corresponding to the extreme cases $X_1=0.1, 1.9, X_2=0.5, 1.5$.

In order to model superspreaders, the contact rate C_1 is assumed to have a Pareto II distribution with shape 2.1, scale 5 and minimum 0. The contact rate after symptoms onset or detection is assumed to be $C_2=\rho C_1$. For example, $\rho=1$ corresponds to no reduction of contact activity after symptoms onset and $\rho=0$ corresponds to complete isolation. Furthermore $a_5=2$, as with this choice the fraction of asymptomatic individuals is around $\frac{1}{4}$, in line with [31]. Monte Carlo integration is used to compute the expectations in (5.4) and (5.6).

6.1. Reducing the contact activity of symptomatic individuals

In this section, we analyse the effect of reducing the contact activity of symptomatic individuals by letting ρ vary between 0 and 1 and reporting variations of the generation time distribution and reproduction numbers in figure 2 and table 2. We recall that the values presented here are not estimates, they are calculated using the formulae of §5.5 with the choice of parameters described at the beginning of this section.

When no interventions are in place, it is still unrealistic to expect no reduction in the contact rate due to symptoms, i.e. $\rho=1$ (first row of table 2 and blue curve in figure 2). In fact, an individual with COVID-19 may naturally reduce contact activity when developing symptoms, even if not explicitly recommended or required to do so. This is simply because symptoms, especially if severe, are often associated with a tendency to reduce social activities, as for example is the case with seasonal flu for which no strict isolation measures are in place. Therefore, the scenario without interventions should correspond to one of the first rows of table 2, for example to $\rho=0.8$, as will be assumed in §6.4. Whereas, the ideal scenario in which all symptomatic individuals are completely isolated at symptoms onset corresponds to $\rho=0$. It is evident that, as expected, isolating symptomatic individuals lowers the mean generation time. We conclude that the variation of the generation time due to reducing the contact activity of symptomatic individuals may be, as in this example, quite significant. While the variation of the reproduction number is also significant,

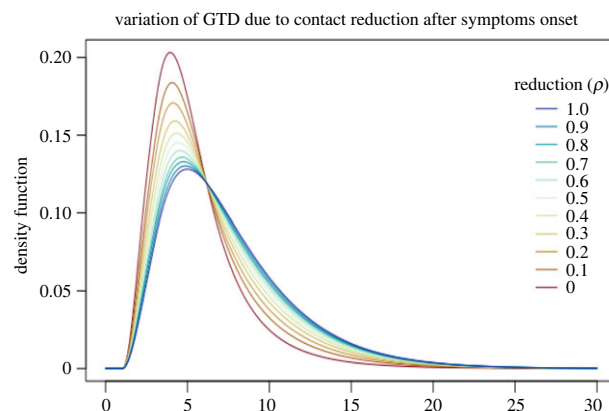


Figure 2. The generation time distribution (GTD) for different values of ρ , the reducing fraction of contact activity at symptoms onset. No other interventions are in place, i.e. $\sigma=0$, $\rho=0$. The blue line corresponds to no reduction ($\rho=1$) and the red line corresponds to complete isolation ($\rho=0$) at symptoms onset.

Table 2. The corresponding reproduction numbers (total, R , before symptoms, $R^{(1)}$, and after symptoms, $R^{(2)}$) and mean generation times (MGT).

ρ	R	$R^{(1)}$	$R^{(2)}$	MGT
1	4.54	1.73	2.81	7.57
0.9	4.26	1.73	2.53	7.48
0.8	3.98	1.73	2.25	7.38
0.7	3.70	1.73	1.97	7.28
0.6	3.42	1.73	1.69	7.15
0.5	3.14	1.73	1.41	6.99
0.4	2.86	1.73	1.13	6.82
0.3	2.57	1.73	0.84	6.59
0.2	2.29	1.73	0.56	6.31
0.1	2.01	1.73	0.28	5.96
0	1.73	1.73	0	5.48

this preventive measure cannot bring the reproduction number below a certain threshold, even when applied perfectly, i.e. $\rho=0$, because of pre-symptomatic and asymptomatic transmission.

6.2. Isolating symptomatic and screened infectious individuals

Starting from the assumptions of the previous subsection with $\rho=0$, we now include screening. That means that when an individual is detected, through symptoms or screening, their contact activity ceases.

In figure 3 and table 3, we report variations of the generation time distribution and of reproduction numbers due to the screening rate σ varying between 0 and 0.1. The impact of uniform screening is moderate; mean generation times and reproduction numbers are not affected as much as in the previous subsection. Furthermore, a rate $\sigma=0.01$ entails a high effort in practice, as it requires roughly 1% of the population being screened each day. Higher values of σ are hardly reachable in practice, and as we mentioned above,

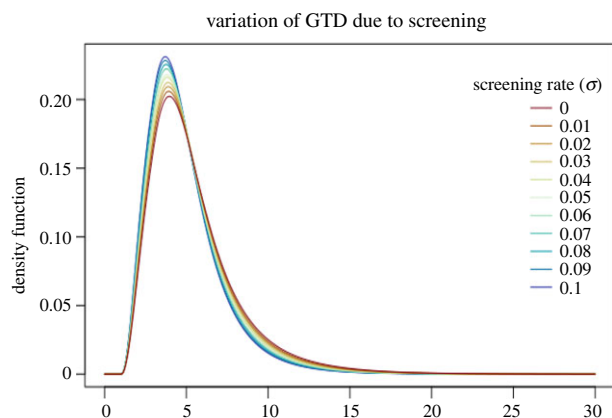


Figure 3. The generation time distribution (GTD) for different values of σ , the screening rate. Detected individuals are completely isolated, i.e. $\rho = 0$, and there is no contact tracing, i.e. $p = 0$.

Table 3. The corresponding reproduction numbers and mean generation times (MGT).

σ	R	MGT
0	1.73	5.48
0.01	1.64	5.40
0.02	1.57	5.31
0.03	1.47	5.25
0.04	1.39	5.20
0.05	1.32	5.13
0.06	1.26	5.07
0.07	1.20	5.00
0.08	1.14	4.94
0.09	1.08	4.90
0.1	1.04	4.84

contact tracing or other targeted testing procedures might be more efficient than increasing the value of σ .

6.3. Isolating symptomatic and contact traced infectious individuals

In this subsection, we exclude screening, i.e. $\sigma = 0$, and we consider contact tracing. At symptoms onset or when they are contact traced, individuals are completely isolated, i.e. $\rho = 0$. The rate at which an individual is detected by through-infectior contact tracing is chosen to be $a_{CT2} = 0.1$. The effect of this rate is analogous to the effect of the screening rate and can thus be seen in the previous subsection. Here instead we focus on illustrating the effect of contact tracing through infectees. To this aim, we let the probability of successful contact tracing, p , vary between 0 and 1, which respectively correspond to the scenario in which no contacts an individual has made can be traced and the scenario in which all contacts can be traced. The results are reported in figure 4 and table 4. The contact tracing delay is chosen to be 24 h, i.e. $d = 1$. This is the time between the detection of an infectee and the detection of their infectior through contact tracing.

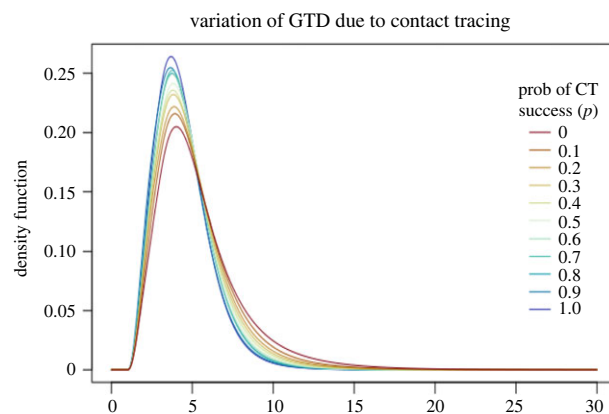


Figure 4. The generation time distribution (GTD) for different values of p , the probability of successful contact tracing. Detected individuals are completely isolated, i.e. $\rho = 0$, and there is no screening, i.e. $\sigma = 0$.

Table 4. The corresponding reproduction numbers and mean generation times (MGT).

p	R	MGT
0	1.73	5.48
0.1	1.50	5.17
0.2	1.45	4.98
0.3	1.31	4.86
0.4	1.21	4.75
0.5	1.16	4.68
0.6	1.10	4.60
0.7	1.05	4.56
0.8	0.94	4.52
0.9	0.90	4.41
1	0.87	4.38

6.4. Biased estimates of reproduction numbers: an example

In the previous subsections, we have illustrated that the generation time distribution may vary significantly due to interventions when symptomatic and detected individuals (through screening or contact tracing) are isolated. In this subsection, we illustrate how this variation may lead to significantly biased estimates of the reproduction number by presenting several scenarios, one without interventions, one with all interventions and others with some interventions. See table 5 for a summary of the chosen parameters for the various scenarios and the corresponding variations.

6.4.1. Neutral scenario: no interventions

Using the same underlying variables as in the previous subsections, we assume that in this neutral scenario without interventions, i.e. $\sigma = 0$, $p = 0$, symptomatic individuals reduce their contact activity by 20%, i.e. $\rho = 0.8$. Imagine that at the beginning of an outbreak, when no preventive measure is in place, a growth rate, $r_0 = 0.23$, and a generation time distribution, g_0 with mean 7.38, are estimated. By the Euler–Lotka equation then an estimate of the basic reproduction number, $R_0 = 3.98$, can be obtained.

Table 5. For each scenario, the table reports the chosen parameters (reducing fraction of contact activity ρ , screening rate σ and fraction of known contacts p) and the resulting growth/decline rate r , mean generation time (MGT) and reproduction numbers R estimated from the actual generation time distribution and \hat{R} estimated from the generation time distribution g_0 of the neutral scenario.

isolation	screening	contact tracing	ρ	σ	p	r	MGT	R	\hat{R}
no	no	no	0.8	0	0	0.23	7.38	3.98	3.98
yes	yes	yes	0.2	0.01	0.7	0.09	6.17	1.67	1.84
yes	no	no	0.2	0	0	0.15	6.31	2.29	2.60
yes	yes	no	0.2	0.01	0	0.14	6.29	2.23	2.53
yes	no	yes	0.2	0	0.7	0.10	6.12	1.73	1.92

6.4.2. All interventions in place

Imagine that, during the same outbreak described in the previous subsection, at a later time, preventive measures are introduced and another growth rate, $r_E = 0.09$, is estimated. We assume that, in this scenario, detected individuals reduce their contact activity by 80%, i.e. $\rho = 0.2$, that roughly 1% of the population is uniformly screened every day, i.e. $\sigma = 0.01$, and that contact tracing is performed in such a way that 70% of the total contacts are reported, i.e. $p = 0.7$.

If we were to use the wide-spread assumption that the generation time distribution does not change during the course of the outbreak, we would use the previously estimated generation time distribution, g_0 , and the Euler–Lotka equation to obtain an estimate of 1.84 for the new reproduction number R_E . However, because of interventions, the generation time distribution has in fact changed from g_0 to g_E , with a mean of 6.17, which is calculated through the formulae in §5.5. Using the actual generation time distribution yields the more precise value of 1.67 for R_E .

Therefore, the variation of the generation time distribution between the neutral scenario and the scenario with interventions causes a non-negligible bias in the estimate of the reproduction number.

The converse problem can also arise if estimates of the generation time distribution obtained under stricter intervention regimes are used to provide estimates of the reproduction number when interventions are less strict or absent. As mentioned in the Introduction, often studies use estimates of the generation time distribution which are based on data collected in China, where extensive contact tracing, including even enforced digital contact tracing, and severe isolation measures were in place. Using this estimate of the generation time distribution to estimate the reproduction number in other countries, where milder preventive measures are in place, may lead to biases, analogously to the situation described above. Likewise, the estimate of the generation time distribution made in one country at a certain time when stricter measures are in place may lead to biases if used in that same country once those measures are eased. For example, it is relevant to take this issue into account when studying the effect of relaxing preventive measures.

The variation of the generation time distribution, and the corresponding bias in the reproduction number estimate, is not always significant and thus can be in some cases neglected; see for example table 5, where various scenarios with various combinations of interventions are reported. In

particular, the closer is the growth/decline rate to zero, the less the estimate of the reproduction number is sensitive to changes of the generation time distribution.

7. Concluding remarks

The contribution of this paper is threefold. A general model, suitable to describe the spread of various infectious diseases under the effect of various preventive measures, is presented. Variations of reproduction numbers and generation times and related biases are analysed. General theoretical formulae are provided and applied in an illustration of a realistic COVID-19 scenario.

The generality of the model relies on the random infectiousness profile and on the random contact activity of infectious individuals, which allows taking inhomogeneity into account. In this way, the infectivity can vary not only over time but also among different individuals, depending for example on the severity of the infection. In §6, we choose a distribution for the infectiousness profile in line with the results in [20,30]; however, the formulae in §5.5 can be used in other settings for any random infectiousness profile, which should be chosen to match appropriate characteristics. Similarly, the random contact activity, which is reduced at a certain random time, allows taking into account different social behaviours of infectious individuals, including for example superspreaders. The model allows a variety of possible choices of explicit forms for distributions, dependencies between quantities and various constants, under non-restrictive modelling assumptions, enabling the representation of most possible interventions.

We perform an analysis of variation of reproduction numbers and generation times. While the variation of the former is a main focus of numerous studies, the variation of the latter is often neglected. In particular, we highlight that variations of the generation time distribution, if neglected, cause bias in the estimate of reproduction numbers. In some scenarios the bias is significant, while in others it is negligible. When doing inference, it is thus crucial to investigate the extent of variations of the generation time distribution in order to determine whether an updated estimate is necessary to avoid significant biases.

While the general formula (3.3) is a valuable theoretical tool for this analysis, we also present realistic, although not fitted to real data, examples in order to illustrate the possible variations and draw attention to the problem. We show that in some cases variations of generation times cause significant

biases in the estimation of reproduction numbers, which are crucial for example to establish control measures. Because of the general nature of our model, the tools we present can be used in numerous scenarios, combined with real data, to investigate the extent of the variation of the generation time distribution. In this way, proper measures can be adopted to avoid biases when needed, that is, either an estimate of the actual generation time distribution is obtained, or variations of generation time distribution are taken into account in the model itself.

Several challenges remain to be solved and interesting new directions are open for future work. In particular, including a more precise model for contact tracing would sharpen the approximation while requiring additional structure to be included in the model. Furthermore, an in-depth study of how individual variation affects the generation time remains to be done, for example how the variation in C_1 , X_1 and X_2 affects the efficacy of different preventive measures. One such feature of relevance would be to analyse which preventive measures are most efficient in detecting superspreaders, which could be modelled by having large variance of C_1 , and in particular, if effective contact tracing is the best method for reducing their effect on an outbreak.

Finally, the random infectivity profiles in this paper are suitable to represent the potentially different infectiousness and severity of symptoms related to different variants of

SARS-CoV-2, thus enabling an analysis of the impact of new variants on reproduction numbers and generation times, while taking into account the effect of preventive measures. A theoretical analysis, which has been tackled in some specific cases (e.g. [32]), would be relevant for the planning of control measures in connection with the emergence of new variants of interest and is left for future work.

Data accessibility. Detailed proofs are available in the electronic supplementary material, supplementary.pdf [33]. The code used for computing generation time distributions and reproduction numbers is available in the file GT.rmd.

Authors' contributions. M.F.: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; G.S.T.: methodology, validation, writing—review and editing; T.B.: conceptualization, funding acquisition, methodology, project administration, supervision, validation, writing—review and editing.

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

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

Funding. T.B. and M.F. are grateful to the Swedish Research Council (grant no. 2020-04744) for financial support. G.S.T. acknowledges the MIUR Excellence Department Project awarded to the Department of Mathematics, University of Rome Tor Vergata (CUP E83C18000100006).

Acknowledgements. The authors thank the anonymous referees for their constructive comments.

References

- Ali ST, Wang L, Lau E, Xu XK, Du Z, Wu Y, Leung GM, Cowling BJ. 2020 Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109. (doi:10.1126/science.abc9004)
- Hart WS, Abbott S, Endo A, Hellewell J, Miller E, Andrews N, Maini PK, Funk S, Thompson RN. 2022 Inference of SARS-CoV-2 generation times using UK household data. *eLife* **11**, e70767. (doi:10.7554/eLife.70767)
- Scarabel F, Pellis L, Ogden NH, Wu J. 2021 A renewal equation model to assess roles and limitations of contact tracing for disease outbreak control. *R. Soc. Open Sci.* **8**, 202091. (doi:10.1098/rsos.202091)
- Torneri A, Libin P, Scalia Tomba G, Faes C, Wood JG, Hens N. 2021 On realized serial and generation intervals given control measures: the COVID-19 pandemic case. *PLoS Comput. Biol.* **17**, e1008892. (doi:10.1371/journal.pcbi.1008892)
- Britton T, Scalia Tomba G. 2019 Estimation in emerging epidemics: biases and remedies. *J. R. Soc. Interface* **16**, 20180670. (doi:10.1098/rsif.2018.0670)
- Champredon D, Dushoff J. 2015 Intrinsic and realized generation intervals in infectious-disease transmission. *Proc. R. Soc. B* **282**, 20152026. (doi:10.1098/rspb.2015.2026)
- Scalia Tomba G, Svensson Å, Asikainen T, Giesecke J. 2010 Some model based considerations on observing generation times for communicable diseases. *Math. Biosci.* **223**, 24–31. (doi:10.1016/j.mbs.2009.10.004)
- Svensson Å. 2007 A note on generation times in epidemic models. *Math. Biosci.* **1**, 300–311. (doi:10.1016/j.mbs.2006.10.010)
- Svensson Å. 2015 The influence of assumptions on generation time distributions in epidemic models. *Math. Biosci.* **270**, 81–89. (doi:10.1016/j.mbs.2015.10.006)
- Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C. 2020 Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936. (doi:10.1126/science.abb6936)
- Flaxman S *et al.* 2020 Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261. (doi:10.1038/s41586-020-2405-7)
- Bi Q *et al.* 2020 Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919. (doi:10.1016/S1473-3099(20)30287-5)
- Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. 2020 Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868. (doi:10.1126/science.abb5793)
- Wallinga J, Teunis P. 2004 Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516. (doi:10.1093/aje/kwh2255)
- Ball F, Donnelly P. 1995 Strong approximations for epidemic models. *Stoch. Process. Their Appl.* **55**, 1–21. (doi:10.1016/0304-4149(94)00034-Q)
- Jagers P. 1975 *Branching processes with biological applications*. New York, NY: John Wiley & Sons.
- Snyder DL, Miller MI. 1991 *Random point processes in time and space*. New York, NY: Springer.
- Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604. (doi:10.1098/rspb.2006.3754)
- Hart WS, Maini PK, Thompson RN. 2022 High infectiousness immediately before COVID-19 symptom onset highlights the importance of continued contact tracing. *eLife* **10**, e65534. (doi:10.7554/eLife.65534)
- He X *et al.* 2020 Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675. (doi:10.1038/s41591-020-0869-5)
- To KK *et al.* 2010 Viral load in patients infected with pandemic H1N1 2009 influenza A virus. *J. Med. Virol.* **1**, 1–7.
- Park SW, Cornforth DM, Dushoff J, Weitz JS. 2020 The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak. *Epidemics* **31**, 100392. (doi:10.1016/j.epidem.2020.100392)
- Ball F, Britton T. 2007 An epidemic model with infector-dependent severity. *Adv. Appl. Probab.* **39**, 949–972. (doi:10.1239/aap/1198177234)

24. Becker N, Starczak DN. 1998 The effect of random vaccine response on the vaccination coverage required to prevent epidemics. *Math. Biosci.* **154**, 117–135. (doi:10.1016/S0025-5564(98)10048-2)
25. Müller J, Kretzschmar M. 2021 Contact tracing—old models and new challenges. *Infect. Dis. Model.* **6**, 222–231. (doi:10.1016/j.idm.2020.12.005)
26. Browne CJ, Gulbudak H, Webb G. 2015 Modeling contact tracing in outbreaks with application to Ebola. *J. Theor. Biol.* **384**, 33–49. (doi:10.1016/j.jtbi.2015.08.004)
27. Browne CJ, Gulbudak H, Macdonald JC. 2022 Differential impacts of contact tracing and lockdowns on outbreak size in COVID-19 model applied to China. *J. Theor. Biol.* **532**, 110919. (doi:10.1016/j.jtbi.2021.110919)
28. Müller J, Kretzschmar M, Dietz K. 2000 Contact tracing in stochastic and deterministic epidemic models. *Math. Biosci.* **164**, 39–64. (doi:10.1016/S0025-5564(99)00061-9)
29. Sun K, Wang W, Gao L, Wang Y, Luo YK, Ren L, Yu H. 2021 Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**, eabe2424. (doi:10.1126/science.abe2424)
30. Hu S *et al.* 2021 Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *Nat. Commun.* **12**, 1533. (doi:10.1038/s41467-021-21710-6)
31. Alene M, Yismaw L, Assemie MA, Ketema DB, Mengist B, Kassie B, Birhan TY. 2021 Magnitude of asymptomatic COVID-19 cases throughout the course of infection: a systematic review and meta-analysis. *PLoS ONE* **16**, e0249090. (doi:10.1371/journal.pone.0249090)
32. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, Thompson RN. 2022 Generation time of the alpha and delta SARS-CoV-2 variants. *Lancet Infect. Dis.* **22**, 603–610. (doi:10.1016/S1473-3099(22)00001-9)
33. Favero M, Scalia Tomba G, Britton T. 2022 Modelling preventive measures and their effect on generation times in emerging epidemics. Figshare. (doi:10.6084/m9.figshare.c.6011452)