



# Effectiveness of isolation measures with app support to contain COVID-19 epidemics: a parametric approach

Andrea Maiorana<sup>1</sup> · Marco Meneghelli<sup>1</sup> · Mario Resnati<sup>1</sup>

Received: 26 August 2020 / Revised: 2 July 2021 / Accepted: 8 September 2021 /  
Published online: 2 October 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

In this study, we analyze the effectiveness of measures aimed at finding and isolating infected individuals to contain epidemics like COVID-19, as the suppression induced over the effective reproduction number. We develop a mathematical model to compute the relative suppression of the effective reproduction number of an epidemic that such measures produce. This outcome is expressed as a function of a small set of parameters that describe the main features of the epidemic and summarize the effectiveness of the isolation measures. In particular, we focus on the impact when a fraction of the population uses a mobile application for epidemic control. Finally, we apply the model to COVID-19, providing several computations as examples, and a link to a public repository to run custom calculations. These computations display in a quantitative manner the importance of recognizing infected individuals from symptoms and contact-tracing information, and isolating them as early as possible. The computations also assess the impact of each variable on the mitigation of the epidemic.

**Keywords** Contact tracing · COVID-19 · Epidemic models

**Mathematics Subject Classification** 92D30 · 62P10

---

✉ Andrea Maiorana  
anm@bendingspoons.com

Marco Meneghelli  
mm@bendingspoons.com

Mario Resnati  
mr@bendingspoons.com

<sup>1</sup> Bending Spoons, Corso Como 15, Milan, Italy

# 1 Introduction

## 1.1 Main concepts and goals

This study aims to develop a probabilistic model to predict the effectiveness of containing an epidemic such as COVID-19 with measures aimed at finding and isolating infected individuals. More precisely, we are interested in modeling such “isolation measures,” by which we mean finding and isolating infected people via their symptoms and contact tracing, to predict the impact of these measures on the effective reproduction number of the epidemic. Special attention is dedicated to the case in which contact tracing is achieved, for a part of the population, through a mobile application.

Studies such as Ferretti et al. (2020) have underlined the role of asymptomatic and presymptomatic transmission in the COVID-19 outbreak, and the consequent importance of using a mobile application for efficient contact tracing. This insight has also led to the development of models to quantitatively assess the impact of a contact tracing app on the epidemic, primarily through agent-based approaches like in Pathogen Dynamics Group (2020).

In this paper, we propose an analytical approach to answer the following questions: How is the effective number  $R_t$  of an epidemic impacted when isolation measures are in place versus when they are not, and what are the main factors contributing to the reduction in  $R_t$ ? We take the effective reproduction number *in the absence* of isolation measures, denoted by  $R_t^0$ , as an input of our model, which is thus independent of any underlying epidemic model. Moreover, our approach is *parametric* in that we concentrate the quantitative description of the isolation measures into relatively few, comprehensible parameters that comprise the input of the model. These parameters include the share of the population using an app, the share of people who self-isolate upon testing positive, and more.

Previous studies concerning the impact on the epidemic of isolating infected individuals include (Müller et al. 2000), which proposes a generative stochastic model of SIR-type, and Fraser et al. (2004), which uses an analytical method more similar to our own. The subject has also been addressed recently in Scarabel et al. (2021) using a deterministic dynamical model.

The starting point of our analysis is the *effective reproduction number*  $R_t^0$  *in the absence of isolation measures*,<sup>1</sup> that we consider as given. When discussing modeling “isolation measures,” we refer to policies focused on selectively isolating infected individuals after these individuals have been found through contact tracing or because they have displayed symptoms. We do not refer to generalized actions like imposing a lockdown, whose impact on the epidemic is considered already known and encompassed in  $R_t^0$ .

$R_t^0$  is defined, for any absolute time  $t$ , as the expected number of cases generated by a random individual who was infected at time  $t$  during their lifetime. This quantity can be written as an integral

<sup>1</sup>  $R_t^0$  must not be confused with the *basic reproduction number*  $R_0$ .

$$R_t^0 = \int_{[0, +\infty)} \beta_t^0(\tau) d\tau,$$

where  $\beta_t^0$  is the *infectiousness* (also called *effective contact rate*):  $\beta_t^0$  is a function describing the expected number of cases generated by an individual infected at time  $t$ , per unit of *infectious age*, that is the period of time (measured in days) elapsed from the time of infection of the individual. So, for example, the number

$$\int_{[1,3)} \beta_t^0(\tau) d\tau$$

is the expected number of people infected between 24 and 72 h from the infector's moment of infection. Note that the normalization  $\beta_t^0/R_t^0$  is the PDF of the *generation time*, the time taken by an individual infected at  $t$  to infect a different individual.<sup>2</sup>

In this study, we set up a methodology and a model to analyze changes in the reproduction number when the population is subject to isolation measures, including the support of an app for individuals who have tested positive, and depending on some parameters of simple interpretation. We denote by

$$R_t = \int_{[0, +\infty)} \beta_t(\tau) d\tau$$

the effective reproduction number *in presence of isolation measures*, and we compute  $R_t$  as a function of  $R_t^0$ ,<sup>3</sup> other epidemiological data such as the symptom onset distribution, and some parameters describing the isolation measures, such as the probability that an infected, symptomatic individual gets a test, or the probability that a recipient of the infection gets notified when their infector receives a positive test. We only model how isolation measures work and how they affect the epidemic,<sup>4</sup> without assuming anything about how the epidemic itself develops. In particular, our model is agnostic of any particular form for  $\beta_t^0$  and  $R_t^0$ .

The final goal of the model we propose is to understand the most important leverages that may facilitate optimization to better direct efforts of decision-makers, scientists, and developers. Such factors include app efficiencies, timeliness of notifications, app adoption in the population, and others.

<sup>2</sup> This is better explained in Sect. A.5.

<sup>3</sup> We stress that the time evolution of  $R_t^0$ , describing how the epidemics would have evolved without the measures we are modeling, is taken as known—our goal is to study the relative impact of the measures. In particular, we do not take into account possible second-order effects on  $R_t^0$ , such as general changes in the behavior of the population, that may come as a consequence of the measures and their impact.

<sup>4</sup> Recall that our model simultaneously includes the effect of isolating infected individuals recognized either through their symptoms or through contact tracing. However, as becomes apparent in the examples of Sect. 4, we will immediately be able to single out the additional impact of contact tracing only, for example.

## 1.2 The assumptions of the model and outline of the paper

The model developed in Sect. 2 is the translation into mathematical terms of the following assumptions, that describe an idealized schema in which infected individuals acknowledge their illness and take measures to avoid infecting others.

- An infected individual who shows symptoms is immediately<sup>5</sup> notified that they should take a test (which does not discount the possibility that they acknowledge this necessity independent of an external input). This process does not always necessarily occur, but does so with a probability  $s^s$ .
- Given a infector–infectee pair, when the infector tests positive after the contagion, the infectee is immediately notified to take a test, with probability  $s^c$ .
- In either scenario, after an infected individual is notified to take a test, they take a test which will return a positive result after a time from the notification, which is distributed according to a given distribution  $\Delta^{A \rightarrow T}$  (possibly reaching  $+\infty$  to account for the case in which the individual is never tested or never receives the positive outcome).
- Immediately upon receiving the positive outcome of the test, an average infected individual will self-isolate with probability  $\xi$ . Put differently, the number of individuals they infect from this moment is reduced by a factor  $1 - \xi$  compared to the scenario in which they do not take any isolation measures.

The equations derived from these hypotheses produce an algorithm that computes the time evolution of the key quantities. This is summarized in Sect. 2.5.

Note that in our model we are only considering *forward* contact tracing, i.e., infectees are notified of the positive result of their infectors, but not vice-versa. Doing otherwise would significantly complicate the discussion. This is probably the main limitation of the model, which may thus underestimate the effectiveness of the isolation measures: While backward contact tracing is in general less effective when timeliness in isolating infected individuals is key, it must be noted that its effect may be significant for epidemics for which super-spreaders, i.e. individuals that infect a large number of people, have a major impact on the contagion. Such individuals may be identified more easily thanks to backward tracing. A treatment of backward tracing in the context of a generative model is covered in Müller et al. (2000, §3.1).

Subsequently, in Sect. 3 we consider a more complex model. Instead, we assume that the population is split into two groups, depending on whether or not they use a mobile application for epidemic control. The parameters  $s^s$  and  $s^c$  are different, depending on whether they refer to individuals who use the app.

Finally, in Sect. 4, we apply these models by computing the suppression of  $R_t$  for specific choices of the input parameters, particularly to assess the importance of such parameters. As for the input parameters that describe the epidemic, we use data relative to COVID-19. All these data are taken for a single source (Ferretti et al. 2020). It should be noted that these quantities are still preliminary, have quite large uncertainties, and

<sup>5</sup> This and the following assumption of immediate notification simplify the treatment to the extent that they avoid adding further distributions modeling some delays. In fact, they are not essential hypotheses, and such real-world delays could be also taken into account in the current setting, by including them in the distribution of the time  $\Delta^{A \rightarrow T}$  between notification and test, introduced below.

are not necessarily the most up-to-date. However, we stress that these data are only used as inputs in all our computations, which can be easily reproduced and extended by using the code available in the open repository (Maiorana and Meneghelli 2021). It would be immediate to redo the computations with different inputs, to reflect any new understandings the scientific community should gain on COVID-19. In addition to this, in Sect. 4.2.3 we briefly check the robustness of our results with respect to changes in some epidemic data, namely the share of infected individuals that are asymptomatic, the contribution of those individuals to the reproduction number, and the generation time distribution.

The paper includes an Appendix where the main steps of the mathematical model are proven rigorously, in a framework where the hypotheses can be formulated precisely using the language of probability theory.

### 1.3 Discussion of the results

Summing up, this paper introduces a model of targeted isolation measures—with special attention paid to those based on contact tracing—in the context of an epidemic with given dynamics. It studies the impact of these, measured as the change in the key indicators of the epidemics (first of all, the reproduction number) with respect to the situation without measures. It presents a methodology to turn the assumptions defining the model into mathematical equations, without assuming an underlying model explaining the time evolution of the epidemic. In particular, the formalism developed in the Appendix allows a careful and exact development of the theory, in which all the interdependencies of the involved quantities are clarified. We end up with a set of equations that express the relevant quantities in terms of those relative to previous times, giving a deterministic time evolution.

These equations (summarized in Sect. 2.5 for the “homogeneous” setting) are quite complex, reflecting the non-triviality of the assumptions about how isolation measures work. This makes it hard to analyze them analytically, for example, to study the asymptotic behaviour of the solutions, as was done in Fraser et al. (2004). On the other hand, our treatment allows us to refrain from making strong and unrealistic independence assumptions about the involved quantities, and leaves us greater freedom in setting up the hypotheses of how contact tracing works (for example, the isolation of contact-traced individuals is not assumed to be certain, nor immediate). And, notably, it allows us to numerically compute, with arbitrary precision, the time evolution of the reproduction number  $R_t$  (and, hence, of the epidemic size) starting from the “default” reproduction number  $R_t^0$ , other epidemiological data, and the parameters introduced in Sect. 1.2 describing the isolation measures.

We stress that, despite our extensive use of the language of probability theory, our model of the isolation measures is deterministic: It works as if the full history of the epidemic, with or without isolation measures, is given, and uses some parameters describing the mean efficacy of the isolation measures on the population. It then expresses  $R_t$  in terms of  $R_t^0$  and these parameters.

Note also that, in this paper, we always refer to  $R_t$  as the *case* reproduction number. Sometimes, the *instantaneous* reproduction number is instead used in the literature

when monitoring the evolution of an epidemic.<sup>6</sup> Our choice is also connected to the way in which we formulate our mitigation hypotheses in Sect. 2 in terms of parameters  $s^s$ ,  $s^c$ ,  $\xi$ , which we consider depending on absolute *infection* times rather than notification and isolation times. An alternative formulation following the latter option would add some slightly more cumbersome formulae but otherwise no essential complications of note to the treatment.

A limitation to the model comes from our homogeneous-mixing hypotheses regarding contact tracing and isolation policies: The only heterogeneity taken into account is the separation between individuals who do or do not use an app in Sect. 3. For example, the fact that, in reality, individuals belonging to the same household are more easily traced (in addition to being more easily infected by each other) is not taken into account. Besides the absence of backward contact tracing, mentioned in Sect. 1.2, other limitations may be attributed to the specific form of the hypotheses. However, many changes to the assumptions could be taken into account within the same mathematical framework: Features such as a different delay in testing for symptomatic or contact-traced individuals, or the existence of a targeted quarantine for potential infected individuals (even before they get tested) could be modeled without adding conceptual complications.

By using the model in Sect. 4 to compute the reduction in  $R_t$ , we can recognize how isolation measures, particularly app-mediated isolation measures, can play an important role in mitigating epidemics like COVID-19. However, our results show how the impact of such measures is strongly sensitive to parameters describing their efficiency and timeliness: For example, the reduction in  $R_t$  quickly becomes insignificant as the time taken to get a positive test result (and then to start isolating) grows past a few days (see Fig. 2).

The computations relative to the case in which an app is used show the importance of having an app which is effective at spotting infections, maximizing the fraction of true-positives.<sup>7</sup> Past studies like Bendavid et al. (2021) and Li et al. (2020) suggest that “standard” contact tracing measures used by healthcare systems may be less efficient (fewer truly infected individuals are recognized) and slower when compared to an app (usually, several days elapse between symptom onset, the first medical visit, and the test outcome). In the computations, we model this fact by setting different parameters for people using an app and people who don’t, with the latter parameters left to reasonably low values. We analyze how the impact on the epidemics depends on these parameters and the app adoption rate (Fig. 9), showing how these are all key factors in reaching satisfactory epidemic mitigation levels.

## 2 The mathematical model in the homogeneous population setting

In this section, we develop the core mathematical model of the paper. We do so with a simplified scenario in which the same isolation measures apply to the entire population,

<sup>6</sup> This typically involves modeling the generation time distribution as constant in time, which is an assumption we do not make. See Cori et al. (2013) for a detailed account of this matter.

<sup>7</sup> To be trusted by its users, the app should also aim at reducing the fraction of false-positives. This is something that our study does not consider.

thus eliminating the need to distinguish between those who do and who do not use an app. Some mathematical derivations require extra care, and their complete proofs have been moved to the Appendix to prevent this section from being loaded with many formulae and a heavier formalism.

## 2.1 Notations and conventions

We consider random variables on the sample space of all infected individuals, describing (absolute) times at which certain events happen:  $t^I$  (time of infection),  $t^S$  (time of symptom onset),  $t^A$  (time of infection notification),  $t^T$  (time of positive test). These variables can take  $+\infty$  as a value to express the cases in which an event never takes place (this is useful when writing relations between them).

As we want to relate these variables to the reproduction number  $R_t$ , which measures the average number of people infected by an individual infected at a given time  $t$ , it is logical that all these variables refer to the infectious age (that is, the time from the infection) of the average individual infected at  $t$ : so we have, for example, the relative time of symptom onset, which is the  $[0, +\infty]$ -valued random variable

$$\tau_t^S = (t^S - t^I)|_{t^I=t} = t^S|_{t^I=t} - t.$$

We can assume that this variable is independent of the contagion time  $t$ . Hence, we denote it by  $\tau^S$ . Analogously, we have the random variables  $\tau_t^A$  (time of notification for an individual infected at  $t$ , measured since  $t$ ),  $\tau_t^T$  (time of positive test for an individual infected at  $t$ , measured since  $t$ ).

In this section we need to understand how to describe the random variables  $\tau^S$ ,  $\tau_t^A$ ,  $\tau_t^T$ , and their relation to the reproduction number

$$R_t = \int_{[0, +\infty)} \beta_t(\tau) d\tau,$$

based on the assumptions of Sect. 1.2. The finite parts of these random variables are described using improper CDFs, denoted by  $F^S$ ,  $F_t^A$ , and  $F_t^T$  respectively, whose limit for  $\tau \rightarrow +\infty$  (representing the probability that each time is less than infinite) may be less than 1. So, for example,  $F_t^T(\tau)$  denotes the probability that an individual infected at  $t$  tests positive within a time  $\tau$  from the time of infection.  $\lim_{\tau \rightarrow +\infty} F_t^T(\tau)$  is the probability that the same individual eventually tests positive.

Further auxiliary variables are introduced later on.

## 2.2 The suppression model for $R_t$

Recall from Sect. 1.2 how we assume that self-isolation works: If an infected individual tests positive, then they immediately self-isolate, resulting in a reduction, on average, of the number of people they subsequently infect by a multiplicative factor  $1 - \xi_t$ ,

which we assume given, and possibly depending on the time  $t$  at which the individual was infected.<sup>8</sup>

We can then determine a relation between the “default” reproduction number density  $\beta_t^0$ , its correction  $\beta_t$  as a result of the isolation measures, and the distribution of the relative time  $\tau_t^T$  at which individuals infected at  $t$  receive a positive test result. This relation holds for any  $t$  greater or equal to the time  $t_0$  at which the isolation measures are enacted.

For simplicity, let’s assume for a moment that receiving a positive test and infecting someone (assuming no isolation measures) at a given infectious age  $\tau$  are independent events. By  $\tau$ , an individual who was infected at  $t$  has already received a test with probability  $P(\tau_t^T < \tau) = F_t^T(\tau)$ . In such a case, the number of people they infect per unit time is  $(1 - \xi_t)\beta_t^0(\tau)$ . Alternatively, if the individual has not received a test by  $\tau$  (which happens with probability  $1 - F_t^T(\tau)$ ), they do not self-isolate, and the average number of people they infect per unit time is just  $\beta_t^0(\tau)$ . In summary, we have, for any  $\tau \in [0, +\infty)$ ,

$$\begin{aligned}\beta_t(\tau) &= F_t^T(\tau)(1 - \xi_t)\beta_t^0(\tau) + (1 - F_t^T(\tau))\beta_t^0(\tau) \\ &= \beta_t^0(\tau) \left(1 - \xi_t F_t^T(\tau)\right).\end{aligned}\quad (1)$$

This is analogous to Eq. 6 in Fraser et al. (2004). To illustrate further, suppose that *all* infected individuals test positive at the same infectious age  $\tau_T$ , i.e.  $F_t^T(\tau)$  is a Heaviside function with step at  $\tau_T$ : then we have  $\beta_t(\tau) = \beta_t^0(\tau)$  for  $\tau < \tau_T$  and  $\beta_t(\tau) = (1 - \xi_t)\beta_t^0(\tau)$  for  $\tau \geq \tau_T$ .

However, the above result relies on the assumption of independence between testing positive and the number of people the individual would infect without isolation. In practice, this is not an adequate reflection of what occurs. For example, with COVID-19, it is known that a significant proportion of the infected population is asymptomatic, and less contagious—see e.g. Mizumoto et al. (2020) and Ferretti et al. (2020). Given the lack of symptoms, this population has a lower probability of self-isolating. To overcome this factor, we introduce a new random variable  $G$ , which has a finite range  $\{g_1, \dots, g_n\}$  that describes the severity of symptoms of an infected individual. It is assumed to be independent of the time  $\tau^S$  of symptom onset, but it is related to the number of infected people and the probability of the individual recognizing their own symptoms. Then, to write a relation between  $F_t^T$  and  $R_t$ , we restrict the relevant random variables to each possible value of  $G$ : for any  $g = g_1, \dots, g_N$  we denote by

<sup>8</sup> Equivalently, this hypothesis could be viewed as the assumption that an individual who tests positive either self-isolates completely, without infecting anyone else from that moment, or, alternatively, does nothing, with the first circumstance happening with probability  $\xi_t$ .



$$F_{t,g}^T(\tau)$$

the probability that an individual infected at time  $t$  and with severity  $g$  has tested positive by  $\tau$ . Similarly, we denote by

$$R_{t,g} = \int_{[0,+\infty)} \beta_{t,g}(\tau) d\tau$$

the average number of people infected by an individual infected at  $t$  and with severity  $g$ , and by  $R_{t,g}^0$  the analogous quantity in absence of isolation measures. Assuming now that for a given  $g$  the number of people infected (without isolation) and the event of being tested are independent, we write our “suppression formula” as

$$\beta_{t,g}(\tau) = \beta_{t,g}^0(\tau) \left(1 - \xi_t F_{t,g}^T(\tau)\right). \quad (2)$$

In Sect. A.3 we include a careful derivation of this formula. Note that the relations with the aggregate variables are

$$F_t^T = \sum_g p_g F_{t,g}^T, \quad R_t = \sum_g p_g R_{t,g},$$

where  $p_g = P(G = g)$  is the probability that an infected individual has symptoms with severity  $g$ .<sup>9</sup>

Also, in Sect. 4 we always take  $G$  to assume the values 0 and 1 only, to describe asymptomatic versus symptomatic infected individuals. However, this formalism allows for a greater diversification of  $R_t^0$ , according to the severity of the illness.

We end this subsection with an example of an application of (2) in a simplified scenario. Suppose that  $G$  only takes the values 0 and 1, describing asymptomatic and symptomatic infected individuals, and that each constitutes half of the population. Suppose also that  $\xi_t = 1$ , and that asymptomatic individuals are never tested, so that  $F_{t,0}^T = 0$ , while symptomatic individuals are tested immediately after infection, so that  $F_{t,1}^T(\tau) = \theta(\tau)$ , where  $\theta$  is the Heaviside function. Then, we have  $R_{t,0} = R_{t,0}^0$  and  $R_{t,1} = 0$ , so that  $R_t = R_{t,0}^0/2$ . Had we used Eq. (1) instead, we would have ended up with  $R_t = R_t^0/2$ , which does not take into account the fact that isolating symptomatic individuals has a greater impact on the reduction of  $R_t$  than isolating the same proportion of randomly chosen individuals.

<sup>9</sup> Note that, according to our convention,  $R_t$  is the *weighted average* of its components  $R_{t,g}$ . Often, in the literature (e.g. in Ferretti et al. 2020) a different convention is used, according to which the components *sum* to  $R_t$ . To switch to the latter convention, each  $R_{t,g}$  should be divided by the respective probability  $p_g$ .

### 2.3 First considerations on the variables $\tau^S$ , $\tau_t^A$ , and $\tau_t^T$

The distribution of the time  $\tau^S$  of symptom onset is independent of the isolation policy and is considered as given throughout the paper, although its specific shape is irrelevant in this section.<sup>10</sup>

The description of  $\tau_t^A$  is addressed in the next subsection. Here, we only consider its relation with  $\tau_t^T$ : Having assumed that the time between notification and testing positive is described by a given random variable  $\Delta^{A \rightarrow T}$ , which is independent from  $\tau_t^A$  and for simplicity constant in absolute time, we have

$$\tau_t^T = \tau_t^A + \Delta^{A \rightarrow T}.$$

The relation still holds if we restrict it to individuals with a given severity  $g$ , and hence

$$F_{t,g}^T(\tau) = \int_{[0,+\infty)} F_{t,g}^A(\tau - \tau') dF^{A \rightarrow T}(\tau'), \tag{3}$$

where  $F^{A \rightarrow T}$  is the improper CDF of  $\Delta^{A \rightarrow T}$ .

### 2.4 Describing $\tau_t^A$

In this subsection, we consider the random variable  $\tau_t^A$  and study the relations with it that formalize the assumptions of Sect. 1.2, namely:

- When an infected individual shows symptoms, they receive an immediate notification to get tested, with probability  $s_{t,g}^S$  depending on the severity  $g$  of symptoms, and possibly on the infection time  $t$ .
- Immediately after an infector tests positive, each infectee is notified of the risk, with probability  $s_t^C$ . If the contagion takes place after the positive test, then the infectee is never notified.

We introduce two new random variables relative to individuals infected at a given time  $t$ , describing the receiving of a notification for either cause:

- We denote by  $\tau_{t,g}^{A,S}$  the time from infection at which an individual infected at  $t$  and with severity  $g$  is notified because of symptoms. We assume that this happens with probability  $s_{t,g}^S$  at the time  $\tau^S$  of the symptom onset, so its improper CDF is simply<sup>11</sup>

$$F_{t,g}^{A,S} = s_{t,g}^S F^S. \tag{4}$$

- We denote by  $\tau_t^{A,C}$  the time from infection at which an individual infected at  $t$  receives a notification resulting from the positive test of their infector. Below, we see how to describe this.

<sup>10</sup> In Sect. 4 we take  $F^S$  to be a log-normal distribution, following the literature.

<sup>11</sup> For simplicity, we use a unique distribution  $F^S$  for all degrees of severity. For asymptomatic individuals,  $s_{t,g}^S$  would be equal to 0.

The relation between these new variables and  $\tau_{t,g}^A$  is

$$\tau_{t,g}^A = \min \left( \tau_{t,g}^{A,s}, \tau_t^{A,c} \right).$$

In terms of improper CDFs, and assuming independence of the two notification times, this gives

$$F_{t,g}^A = F_{t,g}^{A,s} + F_t^{A,c} - F_{t,g}^{A,s} F_t^{A,c}. \tag{5}$$

Describing  $\tau_t^{A,c}$  requires the introduction of an additional random variable  $\tau_t^\sigma$ , that gives, for any individual infected at  $t$ , the time elapsed between the the infection time of their infector and  $t$ . In particular, we need the joint distribution of  $\tau_t^\sigma$  and the severity  $G$ , that can be described in terms of improper CDFs  $F_t^{\sigma,g}$ : Let

$$F_t^{\sigma,g}(\tau)$$

denote the probability that, given an individual infected at  $t$ , their infector has severity  $g$  and was infected at a time  $t' \geq t - \tau$ . Note that these improper CDFs satisfy a normalization condition

$$\lim_{\tau \rightarrow +\infty} \sum_g F_t^{\sigma,g}(\tau) = 1,$$

and they are completely determined by quantities relative to times preceding  $t$ , namely the number of infected people and the infectiousness (more details on how they are computed are deferred to Sect. A.5).

Now, the notification time  $\tau_t^{A,c}$  of an individual infected at  $t$  is by hypothesis equal to the testing time  $\tau_t^T$  of the infector minus the generation time  $\tau_t^\sigma$ , but only if the notification actually occurs, which happens with probability  $s_t^c$  provided that the contagion took place before  $\tau_t^T$ . Hence, to get the improper CDF  $F_t^{A,c}$  we should first average  $F_{t-\tau}^T$ , translated to the left by  $\tau = \tau_t^\sigma$ , over all possible values of  $\tau > 0$ , each weighted by the probability of the generation time being  $\tau$ . In doing this we should also treat separately the different severity levels that the infector may have, as these impact the testing time distribution. So  $F_t^{A,c}(\rho)$  should look like a sum

$$s_t^c \sum_g \int_{(0,+\infty)} F_{t-\tau,g}^T(\rho + \tau) dF_t^{\sigma,g}(\tau).$$

This formula doesn't take into account that by assumption the notification can only occur after the contagion time, meaning that  $F_t^{A,c}$  must be supported on positive numbers. This is considered by replacing the integrand with the probability

$$P \left( \tau < \tau_{t-\tau,g}^T \leq \rho + \tau \right) = F_{t-\tau,g}^T(\rho + \tau) - F_{t-\tau,g}^T(\tau).$$

Also, in averaging the CDFs  $F_{t-\tau}^T$  we should take into account the fact that the testing time of the infector is not distributed like the testing time of an arbitrary individual: Having infected someone at the infectious age  $\tau$ , the infector is more likely than average to be tested *after*  $\tau$ , or to never receive a test. As we will show carefully in the Appendix, to take this into account we need to divide the integrand by the same suppression factor  $1 - \xi_t F_{t,g}^T(\tau)$  that appears in Eq. (2), evaluated at  $t - \tau$ . We conclude that, for any  $\rho \geq 0$ , we have

$$F_t^{A,c}(\rho) = s_t^c \sum_g \int_{(0,+\infty)} \frac{F_{t-\tau,g}^T(\rho + \tau) - F_{t-\tau,g}^T(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g}^T(\tau)} dF_t^{\sigma,g}(\tau). \tag{6}$$

This result is proven rigorously in Sect. A.6.

### 2.5 Summary and discrete-time algorithm

In this section, we have translated the hypotheses made in Sect. 1.2 into mathematical equations describing a dynamical system. In doing this, we added a few natural assumptions of independence between the variables under considerations, namely:

- the assumption in Sect. 2.2 that the testing time of an individual *with given severity* is independent from their *default* infectiousness
- the assumption of independence between notification times  $\tau_{t,g}^{A,s}$  and  $\tau_t^{A,c}$  and the testing delay  $\Delta^{A \rightarrow T}$

Putting all the equations together, we see that we can compute, at any time  $t$ , the suppressed infectiousness  $\beta_{t,g}$  in terms of the parameters  $s_{t,g}^s, s_t^c, \xi_t, \Delta^{A \rightarrow T}$  of the model, the default infectiousness  $\beta_{t,g}^0$  and the other known epidemiological quantities, and the distributions relative to previous times  $t' < t$ .

To add an initial condition to the dynamical system, we assume that the isolation measures start at a given absolute time  $t_0$ , so that  $s_t^s = s_t^c = \xi_t = 0$  for  $t < t_0$ .<sup>12</sup> Hence, all individuals infected at  $t < t_0$  will never take a test (even after  $t_0$ ) and never self-isolate. As a consequence, the effective reproduction number is  $R_t^0$  for  $t < t_0$ , while it gets reduced according to Eq. (2) for  $t \geq t_0$ . In particular, individuals infected at  $t = t_0$  can only be notified of the need to take a test through symptoms, so that  $F_{t_0}^{A,c} = 0$ .

Our set of equations can be approximated with arbitrary precision to a discrete-time algorithm that computes how the epidemic evolves, given the above data.<sup>13</sup> This is the algorithm used in the calculations of Sect. 4.

Summing up, for each time  $t \geq t_0$ , the algorithm works as follows:

1. Compute the number  $v_t$  of individuals infected at  $t$  and the improper CDFs  $F_t^{\sigma,g}$ , from  $v_{t'}$  and  $\beta_{t',g}$  for  $t' < t$  (as detailed in Sect. A.5).

<sup>12</sup> Note that this doesn't prevent us from modeling isolation measures gradually put into place, which can be done simply by taking these parameters to be continuous in  $t$ .

<sup>13</sup> In this discrete setting, all the integrals appearing in the equations reduce to finite sums. In fact, our approach in the Appendix is to derive the same equations starting from a discrete probabilistic model.

2. Compute the distribution of  $\tau_{t,g}^{A,s}$  as in Eq. (4):

$$F_{t,g}^{A,s} = s_{t,g}^s F^S.$$

3. Compute the distribution of  $\tau_t^{A,c}$  from  $F_t^{\sigma,g}$  and the distribution of  $\tau_{t',g}^T$ , for  $t' < t$ , using Eq. (6). If  $t = t_0$ , just take  $F_t^{A,c} = 0$ .
4. Compute the distribution of  $\tau_{t,g}^A$  using Eq. (5), that is

$$F_{t,g}^A = F_{t,g}^{A,s} + F_t^{A,c} - F_{t,g}^{A,s} F_t^{A,c},$$

and then the distribution of  $\tau_{t,g}^T$  from  $F_{t,g}^A$  via Eq. (3).

5. Compute  $\beta_{t,g}$  using the distribution of  $\tau_{t,g}^T$ , via Eq. (2):

$$\beta_{t,g}(\tau) = \beta_{t,g}^0(\tau) \left(1 - \xi_t F_{t,g}^T(\tau)\right).$$

### 3 The extended model including the use of an app for epidemic suppression

So far, we have operated under the hypothesis that the ability to inform infected people that their source has been infected can be described by a single (possibly time-dependent) parameter  $s_t^c$ . Now, let's suppose that the population is divided into people who use an app for epidemic control and people who do not. This forces us to complicate the model of Sect. 2 because, when we analyze the distribution of the notification time  $\tau_t^{A,c}$  for people with the app, we need to apply different weights to the cases in which the source of the contagion has the app or does not. We also leave open the possibility that people using the app may have a different probability  $s_t^s$  of requiring a test because of their symptoms.

The generalization of the homogeneous scenario to this case is quite straightforward. In any case, some more mathematical detail has been added in Sect. A.7.

#### 3.1 Parameters and random variables in the two-component model

A share  $\epsilon_{t,app}$  of the infected population, perhaps depending on the absolute time  $t$ , uses an app that may do the following:

- It gives the users clear instructions on how to behave when they have symptoms indicative of the disease, assuming that this can increase the probability that an infected individual asks the health authorities to be tested because of their symptoms.
- It notifies the users when they have had contact with an infected individual who also uses the app, assuming that this can increase the probability that an infected individual asks the health authorities to be tested because of contact with an infected person.

We then distinguish  $s_{t,g}^s$  into  $s_{t,g}^{s,app}$  and  $s_{t,g}^{s,no\ app}$ , describing the probability that an individual infected at  $t$ , respectively with or without the app, is notified of the need to be tested given that they have symptoms with severity  $g$ . Note that

$$s_{t,g}^s = \epsilon_{t,app} s_{t,g}^{s,app} + (1 - \epsilon_{t,app}) s_{t,g}^{s,no\ app}, \tag{7}$$

so that this distinction does not complicate the model, and is made only for adding clarity in the computations.

The increased complexity of this situation lies in the fact that  $s_t^c$  now has to be replaced by two parameters  $s_t^{c,app}$  and  $s_t^{c,no\ app}$ , describing the probabilities that, given an infector–infectee pair, the positive testing of the infector occurred after the infection caused a notification to be sent to the infectee, respectively in the cases that *both* the infector and the infectee have the app, and that *at least one* of them does not have the app. Note that there is no relation between  $s_t^{c,app}$  and  $s_t^{c,no\ app}$  and the general  $s_t^c$  as simple as Eq. (7).

We also distinguish each random variable between people with the app and people without it. For example, the time of notification due to contact now reads  $\tau_{t,app}^{A,c}$  for people with the app and  $\tau_{t,no\ app}^{A,c}$  for people without it. The relation between their improper CDFs is

$$F_t^{A,c} = \epsilon_{t,app} F_{t,app}^{A,c} + (1 - \epsilon_{t,app}) F_{t,no\ app}^{A,c}.$$

We have analogous formulae for  $\tau_t^T$  and  $\tau_t^{A,s}$ , while there is no need to make a distinction for  $\tau^S$ .

Likewise, we have to separate  $R_t$  into two components  $R_{t,app}$  and  $R_{t,no\ app}$ , namely, the average number of people infected by someone infected at  $t$  who has or does not have the app, respectively:

$$R_t = \epsilon_{t,app} R_{t,app} + (1 - \epsilon_{t,app}) R_{t,no\ app}.$$

Analogous relations hold when restricted to individuals whose illness has a given severity  $g$ .

It is reasonable to assume that having or not having the app is independent of symptom severity, so that, for example, the fraction of individuals infected at time  $t$  using the app and with severity  $g$  is  $\epsilon_{t,app} p_g$ . Also, while of course having an app does impact the testing time distribution and the infectiousness, we can safely suppose that it is independent of the *default* infectiousness, i.e. the number of people an individual would have infected in the absence of measures. This is why in this scenario the suppression formula (2) simply becomes

$$\beta_{t,g,a}(\tau) = \beta_{t,g}^0(\tau) \left(1 - \xi_t F_{t,g,a}^T(\tau)\right), \tag{8}$$

for  $a = \text{app}, \text{no app}$ .

### 3.2 The mathematical relations between the random variables

Now, we can write the new relations between the random variables. Eq. (4) is replaced by

$$F_{t,g,app}^{A,s} = s_{t,g}^{s,app} F^S, \quad F_{t,g,no\ app}^{A,s} = s_{t,g}^{s,no\ app} F^S.$$

The relations (3), (5) immediately extend to each component.

The distributions of  $\tau_{t,app}^{A,c}$  and  $\tau_{t,no\ app}^{A,c}$  can be computed similarly to as we did in Sect. 2.4 for the homogeneous case. But now, for each of them Eq. (6) needs to be split into two parts, accounting for the cases in which the source of the infection has or doesn't have the app:

$$F_{t,app}^{A,c}(\rho) = s_t^{c,app} \sum_g \int_{(0,+\infty)} \frac{F_{t-\tau,g,app}^T(\rho+\tau) - F_{t-\tau,g,app}^T(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g,app}^T(\tau)} dF_t^{\sigma,g,app}(\tau) + s_t^{c,no\ app} \sum_g \int_{(0,+\infty)} \frac{F_{t-\tau,g,no\ app}^T(\rho+\tau) - F_{t-\tau,g,no\ app}^T(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g,no\ app}^T(\tau)} dF_t^{\sigma,g,no\ app}(\tau). \tag{9}$$

For  $F_{t,no\ app}^{A,c}$  we get a similar equation with  $s_t^{c,app}$  replaced by  $s_t^{c,no\ app}$ : In this case, it doesn't matter whether or not the infector has the app. The equation simplifies to a form analogous to Eq. (6), namely

$$F_{t,no\ app}^{A,c}(\rho) = s_t^{c,no\ app} \sum_g \int_{(0,+\infty)} \frac{F_{t-\tau,g}^T(\rho + \tau) - F_{t-\tau,g}^T(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g}^T(\tau)} dF_t^{\sigma,g}(\tau). \tag{10}$$

Again, we refer to the Appendix for a greater mathematical rigor: The last two equations are derived in greater detail in Sect. A.7.

## 4 Scenarios and calculations

In this section, we use the models introduced in Sects. 2 and 3 to numerically compute the suppression of  $R_t$  due to isolation measures in certain scenarios.

The results reported here, as well as new custom calculations, can be obtained by cloning the public Python repository (Maiorana and Meneghelli 2021).

### 4.1 General considerations

Some inputs of the algorithm developed are parameters or distributions describing the features of the epidemic under consideration. In this section, we focus on COVID-19, and we make the following assumptions, taking all the epidemic data from Ferretti et al. (2020) (Table 1, in particular) for convenience:

- The incubation period  $\tau^S$  is distributed according to a log-normal distribution:

$$F^S(\tau) = F_{N_{0,1}}((\log(\tau) - \mu)\sigma)$$

where  $F_{N_{0,1}}$  denotes the CDF of the standard normal distribution. The parameters  $\mu = 1.64$  and  $\sigma = 0.36$  used here imply that the mean incubation period is  $\simeq 5.5$  days.

- The default infectiousness distribution  $\beta_t^0$  is assumed to depend on the absolute time  $t$  only via a global factor, so that

$$\beta_t^0(\tau) = R_t^0 \rho^0(\tau),$$

where  $\rho^0$  (which also represents the default generation time distribution) integrates to 1. It is described by a Weibull distribution with mean 5.00 and variance 3.61:

$$\rho^0(\tau) = \frac{k}{\lambda} \left(\frac{\tau}{\lambda}\right)^{k-1} e^{-(\tau/\lambda)^k}$$

with  $k = 2.855, \lambda = 5.611$ .

- We simplify the severity of symptoms by considering only two levels of severity:  $g = \text{sym}$  and  $g = \text{asy}$ , respectively for symptomatic and asymptomatic individuals. We take asymptomatic individuals as 40%, and we assume that they account for 5% of  $R_t$ .<sup>14</sup> In formulae, this means that the input parameters of our model are

$$p_{\text{sym}} = 0.6, \quad p_{\text{asy}} = 0.4, \\ \beta_{t,\text{sym}}^0 = \frac{0.95}{0.6} R_t^0 \rho^0, \quad \beta_{t,\text{asy}}^0 = \frac{0.05}{0.4} R_t^0 \rho^0.$$

All these assumptions hold throughout the whole section except for Sect. 4.2.3, where we check how the results change using different epidemic data. The other parameters of the model, describing the isolation measures, are selected later.

As Key Performance Indicators (KPIs) describing the effectiveness of the isolation measures, we look at the reduction of  $R_t$  compared to the value  $R_t^0$  it would take in the absence of measures. We call *effectiveness* of the isolation measures the relative reduction in  $R_t^0$ :

$$\text{Eff}_t := 1 - \frac{R_t}{R_t^0}. \tag{11}$$

Thus,  $\text{Eff}_t = 0$  indicates that there is no effect on  $R_t^0$ , while  $\text{Eff}_t = 1$  describes a complete suppression of the contagion. We will see in Sect. 4.2.3 that the dependency of  $\text{Eff}_t$  on the default reproduction number  $R_t^0$  is very weak.<sup>15</sup> As such, attempts to model a realistic profile for  $R_t^0$  have little relevance to our computations, as any choice of  $R_t^0$  leads to almost the same  $\text{Eff}_t$ . Thus, in the rest of this section we simply take

$$R_t^0 = 1.$$

<sup>14</sup> Note that this does not include pre-symptomatic transmission, which is taken into account within the group  $g = \text{sym}$ .

<sup>15</sup> This dependency is due to higher order effects: A higher  $R_t^0$  means that the distribution of  $\tau_t^\sigma$  is more concentrated on small values, and hence the most recent testing time distributions have a greater weight in the time-evolution equation (6).



Another useful KPI is the probability that an individual infected at a certain time  $t$  is eventually found to be positive, namely the limit

$$F_t^T(\infty) := \lim_{\tau \rightarrow +\infty} F_t^T(\tau).$$

In the remainder of this section, we report the results of some selected calculations, considering both the “homogeneous” scenario of Sect. 2 and, in greater detail, the scenario of Sect. 3, in which an app for epidemic control is used. First, we study how the above KPIs evolve in time for certain input parameter choices. Then, we focus on the limits for  $t \rightarrow +\infty$  of these KPIs, i.e., their “stable” values after a sufficient number of iterations, to study how these vary when we change certain input parameters, leaving the others fixed.

## 4.2 Reduction in $R_t$ with homogeneous isolation measures

First, we perform some calculations in the setting of Sect. 2, where the isolation measures are “homogeneous” within the whole population. We recall the parameters that describe this situation, some of which remain fixed in all the calculations:<sup>16</sup>

Parameter	Meaning	Value
$s_{\text{sym}}^S$	Probability that a symptomatic infected individual is notified of the infection because of their symptoms	Not fixed
$s_{\text{asy}}^S$	As above, but for the asymptomatic	0
$\xi$	Probability that someone testing positive self-isolates	Not fixed
$\Delta^{A \rightarrow T}$	Time from notification to positive testing	Constant distribution, whose value is not fixed at this moment
$t_0$	Time at which isolation measures begin	0

Note that assuming that  $\Delta^{A \rightarrow T}$  is a constant random variable means that we are modeling that all individuals notified of the risk test positive, and take the same time to do so. Although unrealistic, this assumption makes little difference to the results. It is made here for simplicity, although it can be easily changed by using a more realistic  $\Delta^{A \rightarrow T}$ , when this datum is available.

### 4.2.1 Time evolution with isolation due to both symptoms and contact-tracing

We now choose the following parameters, describing an optimistic situation, with reasonable efficiencies in spotting infected individuals:

The results are shown in Fig. 1. Note that immediately at  $t = 0$   $R_t$  drops to around 0.92, as half of the symptomatic individuals are notified as soon as they show some symptoms, and they then infect a reduced number of people. Subsequently,

<sup>16</sup> In all the examples these parameters are constant in time. Hence we remove the subscript  $t$  from them.

Parameter	Value
$s_{sym}^s$	0.5
$s^c$	0.7
$\xi$	0.9
$\Delta^{A \rightarrow T}$	2

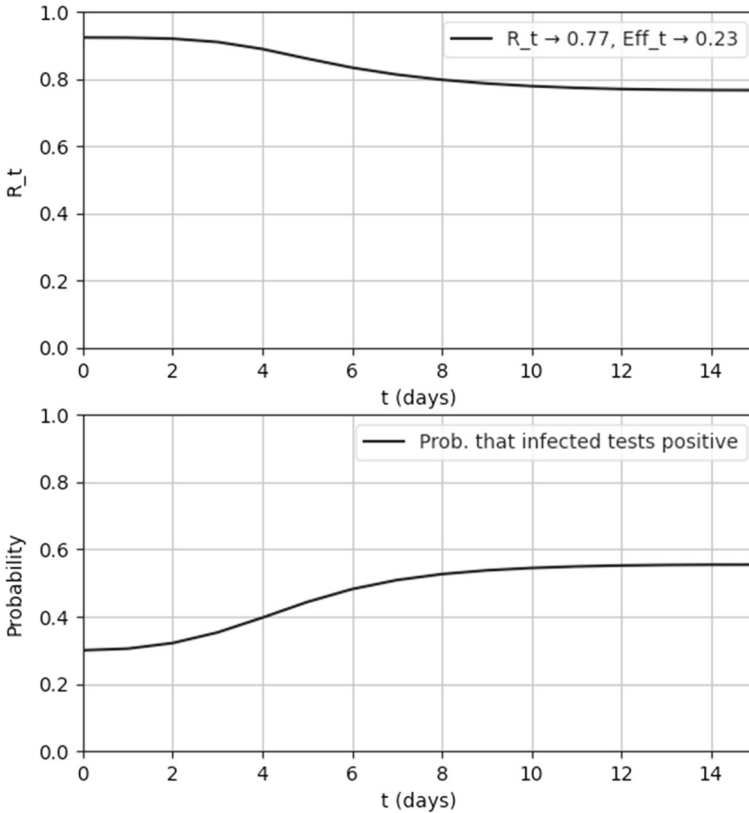


Fig. 1  $R_t$  evolution in the homogeneous model, in an optimistic scenario

$R_t$  continues to decrease due to contact-tracing, quickly approaching its limit value  $R_\infty \simeq 0.77$  (i.e., 84% of the value it would have had with isolation due to symptoms only).

#### 4.2.2 Dependency on testing timeliness

We now focus on the limit value  $Eff_\infty$ , investigating its dependency on the time  $\Delta^{A \rightarrow T}$  from a notification to the positive result of the test (recall that we are assuming that  $\Delta^{A \rightarrow T}$  is a constant random variable).

Like in Sect. 4.2.1, the other parameters are fixed as follows:

Parameter	Value
$s_{sym}^s$	0.5
$s^c$	0.7
$\xi$	0.9

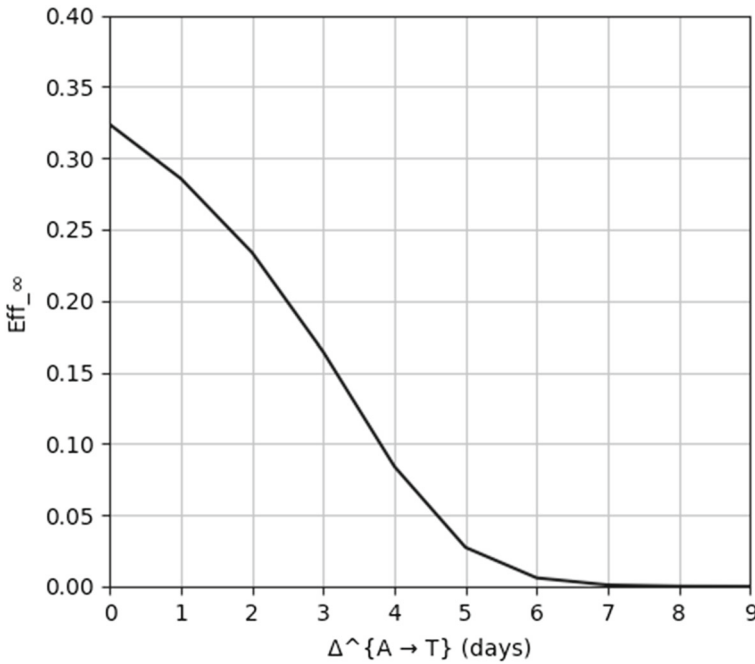


Fig. 2  $Eff_\infty$  as a function of the time  $\Delta^{A \to T}$  from notification to positive testing

The result is plotted in Fig. 2. The effectiveness of the isolation measures improves dramatically with the ability to test (and then isolate) infected individuals as soon as possible after their notification of possible infection.

### 4.2.3 Dependency on the epidemic data used

In this subsection we briefly explore what happens if we change some of the data describing the epidemic, that were introduced in Sect. 4.1 and used elsewhere in this section. This is done to see how  $Eff_\infty$  depends on these data. The other parameters, describing the isolation measures, are fixed as usual:

First, we let the fraction  $p_{sym}$  of symptomatic individuals vary, along with their contribution to  $R_t^0$ —let us denote it here by  $\kappa$ —that is elsewhere taken as  $\kappa = 0.95$ . Recall that

$$R_{t,sym}^0 = \frac{\kappa}{p_{sym}} R_t^0, \quad R_{t,asy}^0 = \frac{1 - \kappa}{1 - p_{sym}} R_t^0.$$

Parameter	Value
$s_{sym}^s$	0.5
$s^c$	0.7
$\xi$	0.9
$\Delta^{A \rightarrow T}$	2

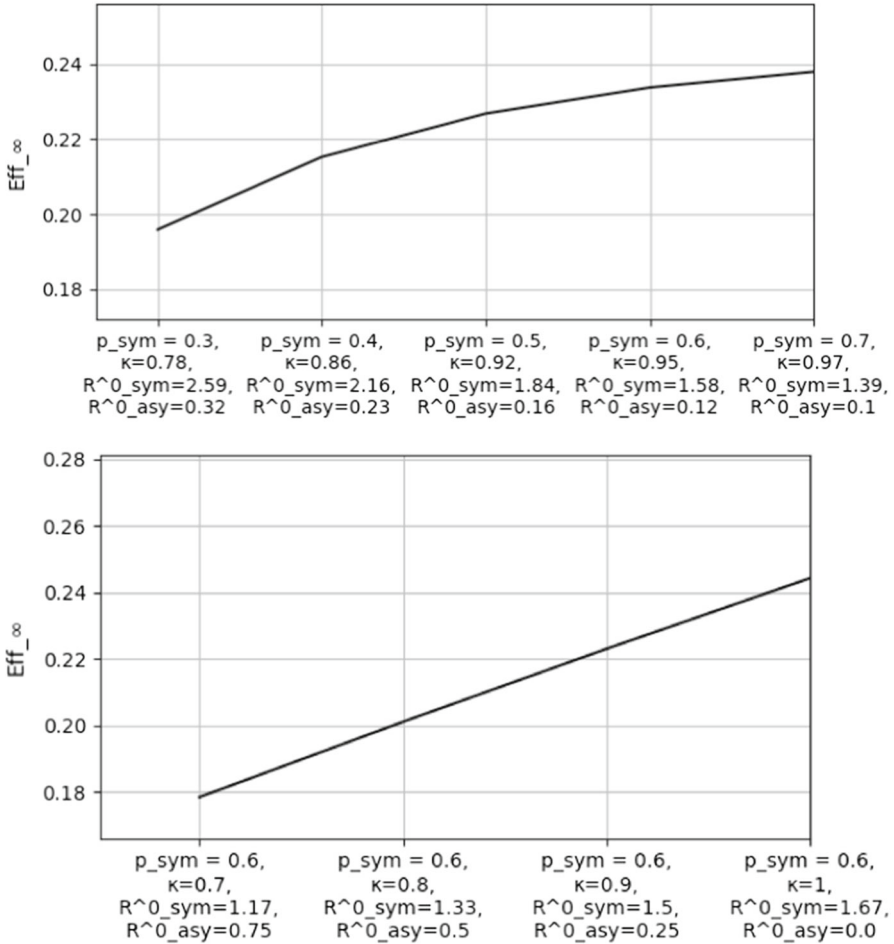


Fig. 3  $Eff_\infty$  for some values of  $p_{sym}$  and  $\kappa$

The value of  $Eff_\infty$  for a few choices of  $p_{sym}$  and  $\kappa$  is plotted in Fig. 3, where it is apparent how the result is robust with respect to changes in these input data. Note that if we fix  $p_{sym}$  and let  $\kappa$  vary, then the two components of  $R_t^0$  (and hence those of  $R_t$ ) are linearly rescaled, meaning that  $Eff_\infty$  also changes linearly.

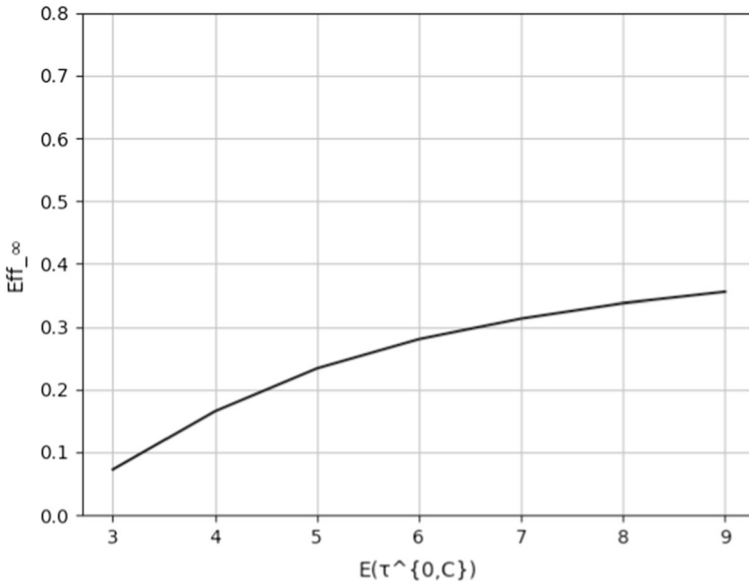


Fig. 4  $Eff_\infty$  for some rescalings of the distribution of  $\tau^{0,C}$

Second, we fix  $p_{sym} = 0.6$  and  $\kappa = 0.95$  as usual, and we modify instead the density  $\rho^0$  of the default generation time, by replacing it with

$$\rho_f^0(\tau) = \frac{1}{f} \rho^0(\tau/f)$$

for  $f > 0$ . Note that this implies that the expected value of the default generation time (denoted here by  $\tau^{0,C}$ ) is multiplied by  $f$ :

$$\mathbb{E}(\tau^{0,C}) = f(5 \text{ days}) .$$

Figure 4 depicts the relation between  $\mathbb{E}(\tau^{0,C})$  and  $Eff_\infty$ , as  $f$  varies. As expected, the isolation measures become more effective as the time taken by the infection to be transmitted increases.

Finally, Fig. 5 shows how  $Eff_\infty$  changes slightly as we change the value of  $R_t^0$  (for  $t \geq 0$ ).

### 4.3 Reduction in $R_t$ in the case of app usage

Now, we focus on applying the model of Sect. 3 to study how  $R_t$  is reduced when a fraction of the population uses an app for epidemic control.

In this case, we summarize the input parameters in the following table, fixing some of them to the given values for the rest of the section (unless explicitly mentioned).

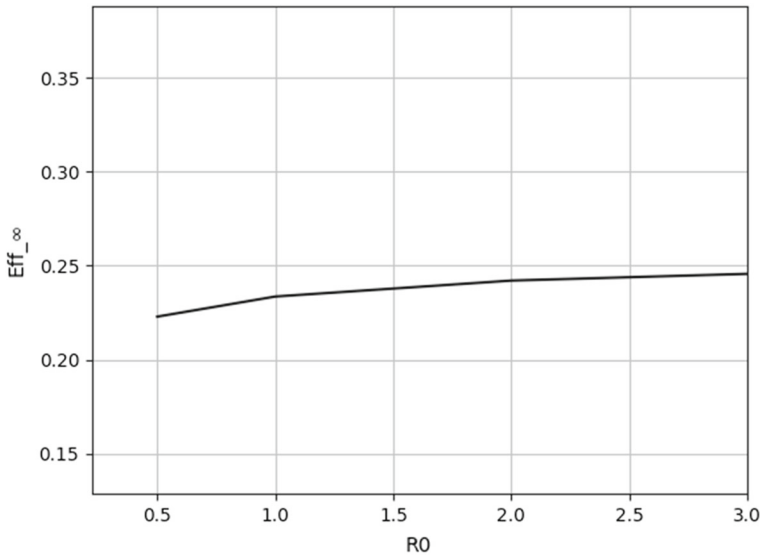


Fig. 5  $Eff_\infty$  for some values of  $R_t^0$ , for  $t \geq 0$

Parameter	Meaning	Value
$s_{sym}^{app}$	Probability that a symptomatic infected individual using the app is notified of the infection because of their symptoms	Not fixed
$s_{sym}^{no app}$	As above, but for individuals without the app	0.2
$s_{asy}^{app}$ , $s_{asy}^{no app}$	As with the two parameters above, but for asymptomatic individuals	0, 0
$s_{c,app}^{asy}$	Probability that an infected individual with the app is notified of the infection because of their source having tested positive	Not fixed
$s_{c,no app}$	Probability that an infected individual without the app is notified of the infection because of their source having tested positive	0.2
$\xi$	Probability that someone testing positive self-isolates	Not fixed
$\Delta^{A \rightarrow T, app}$ , $\Delta^{A \rightarrow T, no app}$	Time from notification to positive testing for people with and without the app, respectively	Constant distributions, whose values are not fixed at this moment
$\epsilon_{t, app}$	Fraction of the population adopting the app at time $t$	Not fixed
$t_0$	Time at which isolation measures begin	0

### 4.3.1 Time evolution in an optimistic scenario

We start with an optimistic scenario, where the app is effective at recognizing infected individuals from symptoms and contact-tracing information. The internal predictive models that estimate the probability of an individual being infected have high efficiencies (a situation likely bound to the possibility of training the predictive models on

real data, in practice). The app is adopted by a large fraction (60%) of the population, and is trusted, so that most of the people notified take a test and self-isolate. The app also helps a notified individual to get tested more quickly.<sup>17</sup>

Parameter	Value
$s_{sym}^{s,app}$	0.8
$s^{c,app}$	0.8
$\xi$	0.9
$\epsilon^{app}$	0.6
$\Delta^{A \rightarrow T, app}$	2
$\Delta^{A \rightarrow T, no app}$	4

As we start from  $R_t^0 = 1$ , we reach a limit value of  $R_\infty \simeq 0.84$  for an effectiveness of 0.16. Note also that  $R_{\infty, app} \simeq 0.75$ , while  $R_{\infty, no app} \simeq 0.97$ . The time evolution of  $R_t$ , along with the other main quantities of interest, is shown in Fig. 6.

### 4.3.2 Time evolution in a pessimistic scenario

We now run an analogous computation in a ‘‘pessimistic’’ scenario. The app can only recognize infected individuals from contact-tracing information, and not from symptoms ( $s_{sym}^{s,app}$  consequently defaults to the no-app value). In addition, we assume a low efficiency  $s^{c,app} = 0.5$ , perhaps due to poor predictive models. Also, only 70% of those testing positive self-isolate.

Parameter	Value
$s_{sym}^{s,app}$	0.2
$s^{c,app}$	0.5
$\xi$	0.7
$\epsilon^{app}$	0.6
$\Delta^{A \rightarrow T, app}$	2
$\Delta^{A \rightarrow T, no app}$	4

Even with a high app adoption rate (60% of the population), the effectiveness drops dramatically. We get  $R_\infty \simeq 0.96$  and  $Eff_\infty \simeq 0.06$ . Most notably, the app does not

<sup>17</sup> Studies such as Li et al. (2020) report that the time from symptom onset to testing through the ‘‘conventional’’ channels (health care system) is in the order of several days. An app is expected to have substantial chances to improve this performance, being a prompt-instrument by construction (for example, when compared with the friction of calling a doctor, inserting symptom descriptions into the app is likely easier), so that

$$\Delta^{A \rightarrow T, app} < \Delta^{A \rightarrow T, no app} .$$

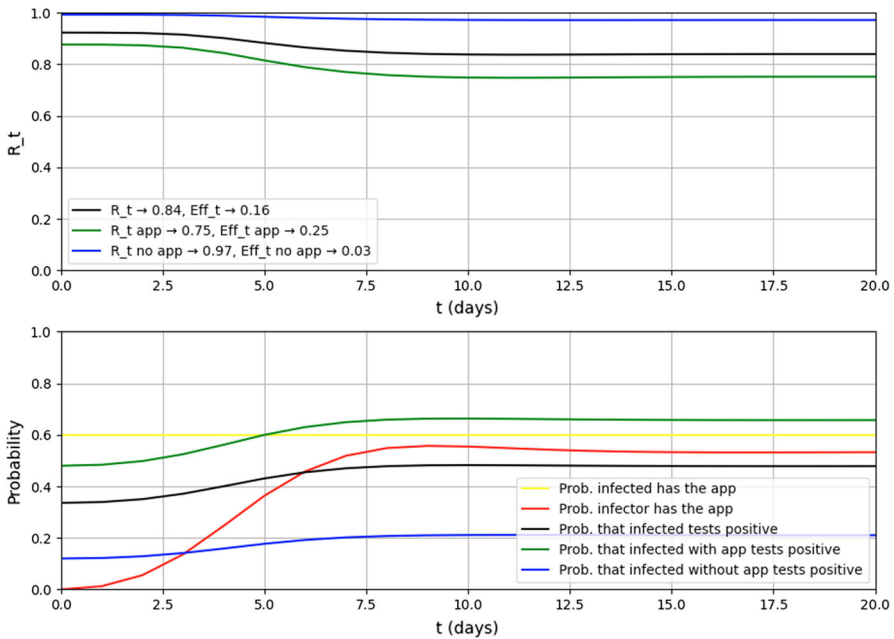


Fig. 6 KPIs evolution in the optimistic scenario

change things much with respect to “standard” isolation measures:  $R_{\infty,app} \simeq 0.95$  and  $R_{\infty,no\ app} \simeq 0.99$ .

### 4.3.3 Time evolution in the case of gradual adoption of the app

Now, we study the evolution of  $R_t$  in a scenario whereby the fraction  $\epsilon_{t,app}$  of people using the app is not constant, but increasing in a linear fashion until it reaches 60% in 30 days:

$$\epsilon_{t,app} = 0.6t/30 \text{ for } 0 \leq t < 30, \quad \epsilon_{t,app} = 0.6 \text{ for } t \geq 30.$$

The other parameters are chosen as in the optimistic scenario of Sect. 4.3.1:

Parameter	Value
$s_{sym}$	0.8
$s_{c,app}$	0.8
$\xi$	0.9
$\Delta A \rightarrow T, app$	2
$\Delta A \rightarrow T, no\ app$	4



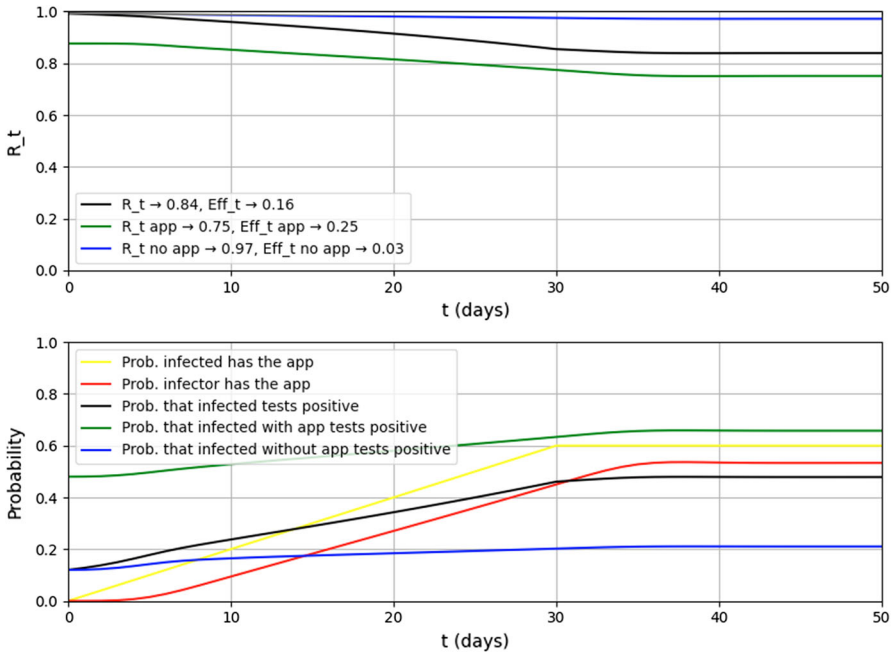


Fig. 7 KPIs evolution in case of gradual adoption of the app

As shown in Fig. 7,  $R_t$  decreases until stabilizing again to the same value obtained in Sect. 4.3.1, although it takes more time to do so. The limit values of the KPIs are not changed by a gradual adoption of the app, compared with a prompt adoption.

#### 4.3.4 Dependency of effectiveness on the efficiencies $s^s$ and $s^c$

We now focus on the study of how the limit values of the KPIs change when we vary certain parameters, starting with the app efficiencies  $s_{\text{sym}}^{s,\text{app}}$  and  $s^{c,\text{app}}$ . In Fig. 8, we plot  $\text{Eff}_\infty$  as a function of these two parameters, while the others are fixed to the following values:

Parameter	Value
$\xi$	0.9
$\epsilon^{\text{app}}$	0.6
$\Delta^{A \rightarrow T, \text{app}}$	2
$\Delta^{A \rightarrow T, \text{no app}}$	4

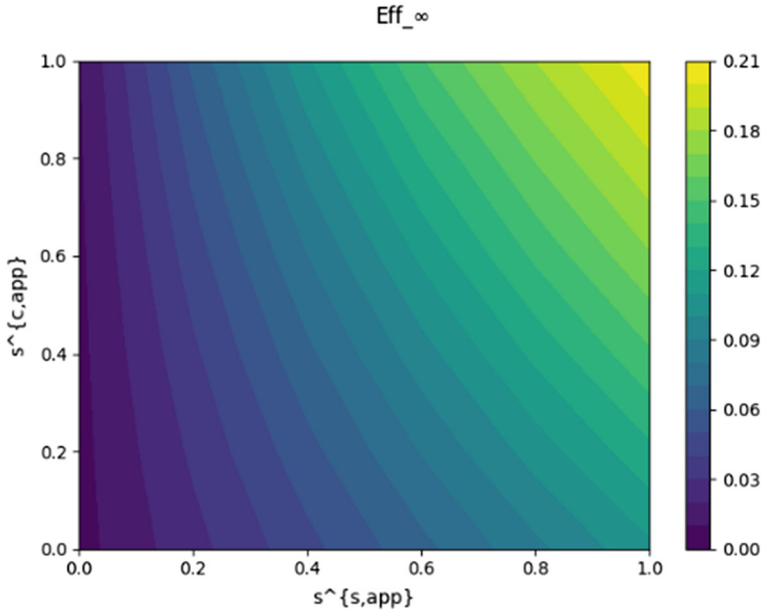


Fig. 8  $Eff_\infty$  as a function of the efficiencies  $s^{s,app}$  and  $s^{c,app}$

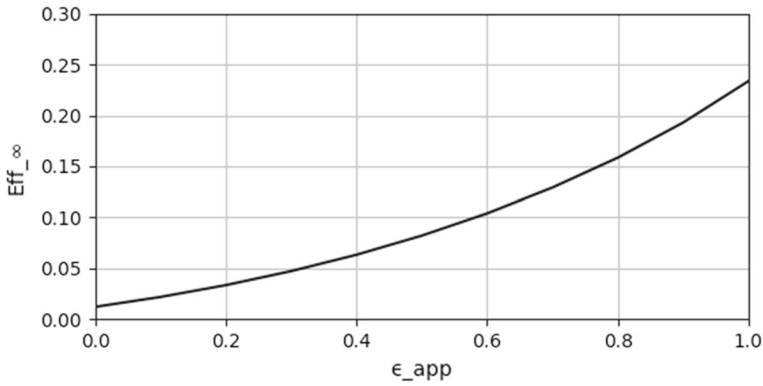


Fig. 9  $Eff_\infty$  as a function of app adoption  $\epsilon_{app}$

### 4.3.5 Dependency on the app adoption

In Fig. 9, we can observe the dependency of the effectiveness  $Eff_\infty$  on the share  $\epsilon_{app}$  of the population using the app. The remaining parameters are fixed to these values:

Parameter	Value
$s_{sym}$	0.5
$s_{c,app}$	0.7
$\xi$	0.9
$\Delta^{A \rightarrow T, app}$	2
$\Delta^{A \rightarrow T, no app}$	4

**Acknowledgements** We thank our colleagues Christy Keenan and Luca Ferrari for carefully proofing the manuscript and providing significant stylistic improvements. We are also grateful to Giorgio Guzzetta and the anonymous referees for several useful comments and suggestions about this study and its publication.

**Author Contributions** M.M. and M.R. conceived the study and its main ideas; A.M. developed the mathematical model; A.M. and M.M. developed the Python repository with the calculations and wrote the manuscript.

## Declarations

**Conflict of Interest** This work was supported by the authors' employer, Bending Spoons S.p.A, which was involved in the development of the contact tracing app adopted by the Italian Government.

## A Appendix: formalism and detailed derivations of mathematical results

The goal of this Appendix is to introduce a mathematical framework in which the hypotheses of our model can be formulated precisely and their consequences proven rigorously. In particular, we will derive the formula (2) describing the suppression of  $R_t$  due to the testing and isolation policies, and the time evolution equation (6). Some more details about the two-component scenario of Sect. 3 are added.

### A.1 Modeling a deterministic epidemics with a probability space

In this subsection, we define the fundamental components of our framework.  $\Omega$  denotes the set of all the individuals infected during the epidemic. It is endowed with two functions. First,

$$t^I : \Omega \rightarrow [0, +\infty)$$

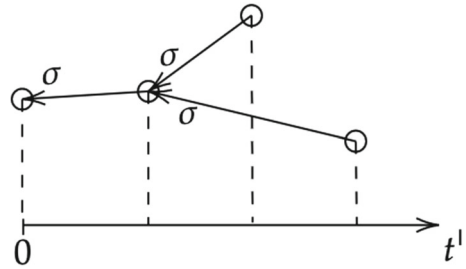
associates to each individual  $\omega \in \Omega$  the absolute time of their infection. Note that we take 0 as the initial time of the epidemics.  $t^I$  partitions  $\Omega$  into a foliation

$$\Omega = \bigcup_{t \in \mathbb{R}} \Omega_t := \bigcup_{t \in \mathbb{R}} \{\omega \in \Omega \mid t^I(\omega) = t\}.$$

We also write

$$\Omega_{>0} := \bigcup_{t \in \mathbb{R}^+} \Omega_t$$

**Fig. 10** Schematic representation of  $\Omega$ ,  $t^I$ , and  $\sigma$ . Dots represent individuals  $\omega \in \Omega$



for the set of individuals infected at a positive time.

Second, for any individual  $\omega \in \Omega_{>0}$  we denote by  $\sigma(\omega) \in \Omega$  their infector. This defines a map

$$\sigma : \Omega_{>0} \rightarrow \Omega.$$

It is natural to assume that, for all  $\omega \in \Omega_{>0}$ ,

$$t^I(\sigma(\omega)) < t^I(\omega).$$

As is the case in the rest of the paper, we will consider other quantities referring to infected individuals and study the mathematical relations between them, with special attention to their *average* properties over all individuals infected at a given time  $t$ . Hence, we will study functions defined over  $\Omega$ , and to talk about their averages over each set  $\Omega_t$  we will introduce a probability measure  $P$  on  $\Omega$ . Since  $\Omega$  is finite and we want to weight all individuals equally,  $P$  is the uniform discrete probability measure:

$$P(E) = \frac{\#E}{\#\Omega},$$

where  $\#E$  denotes the cardinality of a set  $E \subset \Omega$ . Relevant quantities then become random variables, and we are interested in studying their distributions. This always reduces to solving certain counting problems, and introducing  $P$  is largely a way to conveniently write formulas using the language of probability theory. Let us stress that our methodology does not involve any simulation of random processes: The history of the epidemic is completely determined by the triple  $(\Omega, t^I, \sigma)$ , and our study of its evolution consists of writing deterministic relations that express random variables restricted to a time slice  $\Omega_t$  in terms of random variables restricted to slices  $\Omega_{t'}$  for times  $t' < t$ .

The probability measure  $P$  “disintegrates” along  $t_I$ , giving a uniform probability measure  $P_t$  on each  $\Omega_t$ . Also, for any random variable  $X : \Omega \rightarrow \mathbb{R}$  we will denote by

$$X_t := X \upharpoonright_{\Omega_t}$$

its restriction to  $\Omega_t$ . We will be mostly interested in studying the distributions of such restrictions of random variables. In considering quantities like the expected value

$$\mathbb{E}_{P_t}(X_t) = \int_{\Omega_t} X \, dP_t = \frac{1}{\#\Omega_t} \sum_{\omega \in \Omega_t} X(\omega)$$

of  $X_t$ , we will always make the relevant probability measure  $P_t$  explicit to avoid confusion.

### A.2 Infector–infectee pairs, generation time and the reproduction number

Let us introduce some additional notation, for future convenience: First, we define the set

$$\tilde{\Omega} = \{(\omega', \omega) \in \Omega \times \Omega_{>0} \mid \sigma(\omega) = \omega'\}$$

of infector–infectee pairs. This is the graph of the map  $\sigma$ , which thus determines a bijection  $\Omega_{>0} \rightarrow \tilde{\Omega}$ . We also consider two functions describing the generation time: a function  $\tau^C : \tilde{\Omega} \rightarrow \mathbb{R}^+$ , given by

$$\tau^C(\omega', \omega) := t^I(\omega) - t^I(\omega'),$$

and a function  $\tau^\sigma : \Omega_{>0} \rightarrow \mathbb{R}^+$  given by

$$\tau^\sigma(\omega) := \tau^C(\sigma(\omega)).$$

Consider, for all  $\tau \in \mathbb{R}^+$ , the random variable  $n^\tau : \Omega \rightarrow \mathbb{N}$  such that  $n^\tau(\omega)$  is the number of people infected by  $\omega$  at  $\omega$ 's infectious age  $\tau$  (that is, at the absolute time  $t^I(\omega) + \tau$ ):

$$n^\tau(\omega) := \#\{\omega' \in \Omega_{>0} \mid \sigma(\omega') = \omega, \tau^\sigma(\omega') = \tau\}.$$

Also, for  $\tau \in [0, +\infty]$ , we denote by  $N^\tau : \Omega \rightarrow \mathbb{N}$  the random variable that counts all individuals infected *within* the infectious age  $\tau$ :

$$N^\tau(\omega) := \sum_{\tau' \leq \tau} n^{\tau'}(\omega) = \#\{\omega' \in \Omega_{>0} \mid \sigma(\omega') = \omega, \tau^\sigma(\omega') \leq \tau\}.$$

In particular,  $N^\infty(\omega)$  is the total number of people infected by  $\omega$ . The average values of these variables are key indicators of the speed of propagation of the epidemics. In particular, restricting to an absolute time  $t$ , we let

$$R_t := \mathbb{E}_{P_t}(N_t^\infty)$$

be the *effective reproduction number* at  $t$ . Note that averaging on *all* infected individuals simply gives  $\mathbb{E}_P(N^\infty) = 1$ . We also consider the average values of  $N_t^\tau$ , for finite  $\tau$ :

$$B_t(\tau) := \mathbb{E}_{P_t}(N_t^\tau).$$

Notice that  $B_t$  is an improper CDF supported on  $\mathbb{R}^+$ , and because of the finiteness of  $\Omega_t$ , it is a sum of finitely many step functions. It represents the *cumulative infectiousness* of individuals infected at  $t$ , and tends to  $R_t$  for  $\tau \rightarrow +\infty$ .

For practical reasons, it is common to tacitly consider a *continuum limit* in which each  $\#\Omega_t$  tends to infinity and all random variables become continuous.  $B_t$  is then approximated by a smooth function, whose derivative is denoted by  $\beta_t$  (as is the case in the rest of the paper):

$$B_t(\tau) \simeq \int_0^\tau \beta_t(\tau') d\tau',$$

However, in this Appendix we always work in the discrete setting discussed so far, and then we consider the continuum limit only to get formulas for  $\beta_t$ , for consistency with the standard terminology and notation. Using the formalism of measure theory, or simply writing the relations between random variables in terms of their CDFs, allows us to treat both the discrete scenario and the continuum limit in a unified notation.

### A.3 The suppression formula

Here, we discuss in greater mathematical detail the content of Sect. 2.2. In particular, we will derive the suppression formula (Eq. 2) relating the reproduction number  $R_t$  and the distribution of the random variable  $\tau^T: \Omega \rightarrow [0, +\infty]$  describing the infectious age (possibly infinite) at which each individual is tested positive.

To do this, we introduce two additional random variables  $n^{0,\tau}, N^{0,\tau}: \Omega \rightarrow \mathbb{N}$  which are analogues to  $n^\tau$  and  $N^\tau$ , but which instead count the number of individuals that each  $\omega \in \Omega$  would have infected without isolation measures. Similarly, we denote by

$$B_t^0(\tau) := \mathbb{E}_{P_t}(N_t^{0,\tau}), \quad R_t^0 := \mathbb{E}_{P_t}(N_t^{0,\infty}),$$

and  $\beta_t^0$  the analogues of  $B_t, R_t$  and  $\beta_t$  in the absence of isolation measures.

Recall that we assumed the average number of people infected by each individual  $\omega$  is reduced by a factor  $1 - \xi_t$  (possibly depending on the infection time  $t := t^I(\omega)$ ) at times  $\tau$  greater or equal than the testing time  $\tau^T(\omega)$ . This is encoded by the following relation between the expected values of  $n_t^\tau$  and  $n_t^{0,\tau}$  conditioned by  $\tau_t^T$ : For all  $\tau, \rho \in$

$[0, +\infty]$ , we postulate that

$$\begin{aligned} \mathbb{E}_{P_t}(n_t^\tau | \tau_t^T = \rho) &= (1 - \xi_t \delta_{\tau \geq \rho}) \mathbb{E}_{P_t}(n_t^{0,\tau} | \tau_t^T = \rho) \\ &= \begin{cases} \mathbb{E}_{P_t}(n_t^{0,\tau} | \tau_t^T = \rho) & \text{for } \tau < \rho \\ (1 - \xi_t) \mathbb{E}_{P_t}(n_t^{0,\tau} | \tau_t^T = \rho) & \text{for } \tau \geq \rho \end{cases}, \end{aligned} \tag{12}$$

where  $\delta_{\tau \geq \rho} := \chi_{[\rho, +\infty]}(\tau)$  is 1 when  $\tau \geq \rho$  and 0 otherwise. Now we would like to remove the conditioning on  $\tau_t^T$  from the expected values of  $n_t^{0,\tau}$  to get an expression in terms of known quantities only. If, for simplicity, we supposed that  $\tau^T$  and  $n_t^{0,\tau}$  are independent, then Eq. (12) would reduce to  $(1 - \xi_t \delta_{\tau \geq \rho}) \mathbb{E}_{P_t}(n_t^{0,\tau})$ . But, as discussed in Sect. 2.2, this is not a realistic hypothesis. Instead, we only assume that  $\tau^T$  and  $n_t^{0,\tau}$  are independent when restricted to individuals having the same *severity of illness*, which we describe through a random variable

$$G : \Omega \rightarrow \mathbb{R}.$$

In other words, we take  $\tau_t^T$  and  $n_t^{0,\tau}$  to be *conditionally* independent with respect to  $G$ .<sup>18</sup> We assume now that the suppression formula applies equally to individuals of all degrees of severity:

$$\mathbb{E}_{P_{t,g}}(n_{t,g}^\tau | \tau_{t,g}^T = \rho) = (1 - \xi_t \delta_{\tau \geq \rho}) \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau} | \tau_{t,g}^T = \rho), \tag{13}$$

having introduced

$$\Omega_{t,g} := \{\omega \in \Omega_t \mid G(\omega) = g\}$$

and the obvious notation for restrictions of random variables to  $\Omega_{t,g}$  and for the uniform probability measure  $P_{t,g}$  on it. The assumption of conditional independence implies

$$\mathbb{E}_{P_{t,g}}(n_{t,g}^\tau | \tau_{t,g}^T = \rho) = (1 - \xi_t \delta_{\tau \geq \rho}) \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau}). \tag{14}$$

Summing over  $\rho \in [0, +\infty]$  we find

$$\begin{aligned} \mathbb{E}_{P_{t,g}}(n_{t,g}^\tau) &= \sum_{\rho \in [0, +\infty]} P_{t,g}(\tau_{t,g}^T = \rho) \mathbb{E}_{P_{t,g}}(n_{t,g}^\tau | \tau_{t,g}^T = \rho) \\ &= \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau}) \sum_{\rho \in [0, +\infty]} (1 - \xi_t \delta_{\tau \geq \rho}) P_{t,g}(\tau_{t,g}^T = \rho) \\ &= \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau}) (1 - \xi_t F_{t,g}^T(\tau)), \end{aligned}$$

having denoted the improper CDF of  $\tau_{t,g}^T$  by  $F_{t,g}^T$ , as usual.

Finally, to average over all  $\Omega_t$ , we simply notice that

$$\mathbb{E}_{P_t}(n_t^\tau) = \sum_g p_{t,g} \mathbb{E}_{P_{t,g}}(n_{t,g}^\tau),$$

<sup>18</sup> Remember that in this work, the joint distribution of  $n_t^{0,\tau}$  and  $G$  is assumed known, and we want to study how our assumptions on the isolation measures determine the distribution of  $n_t^\tau$ , and hence  $R_t$ .

where  $p_{t,g} := P_t(G_t = g)$ .<sup>19</sup> On the other hand, summing over  $\tau$  gives a relation between  $B_t$  and  $B_{t,g}^0 = \mathbb{E}_{P_{t,g}}(N_{t,g}^{0,\tau})$ :

$$\begin{aligned} B_t(\tau) &= \sum_g p_{t,g} \sum_{\tau' \leq \tau} \mathbb{E}_{P_{t,g}}(n_{t,g}^\tau) \\ &= \sum_g p_{t,g} \int_{(0,\tau]} (1 - \xi_t F_{t,g}^T(\tau')) \, dB_{t,g}^0(\tau'). \end{aligned}$$

Then, we can sum up the content of this subsection as follows:

**Proposition 1** *Take  $t \in [0, +\infty)$ . Assuming the suppression hypothesis (13) and the conditional independence of  $\tau^T$  and  $n_i^{0,\tau}$  with respect to  $G$ , we have*

$$\begin{aligned} \mathbb{E}_{P_t}(n_t^\tau) &= \sum_g p_{t,g} \mathbb{E}_{P_{t,g}}(n_{t,g}^\tau) \\ &= \sum_g p_{t,g} \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau})(1 - \xi_t F_{t,g}^T(\tau)), \end{aligned} \tag{15}$$

where  $p_{t,g} := P_t(G_t = g)$ . Moreover,

$$R_t = \sum_{g \in \mathbb{R}} p_{t,g} R_{t,g} = \sum_g p_{t,g} \int_{\mathbb{R}^+} (1 - \xi_t F_{t,g}^T(\tau)) \, dB_{t,g}^0(\tau).$$

Note that taking the continuum limit of Eq. (15) we retrieve Eq. (2):

$$\beta_t(\tau) = \sum_g p_{t,g} \beta_{t,g}(\tau) = \sum_g p_{t,g} (1 - \xi_t F_{t,g}^T(\tau)) \beta_{t,g}^0(\tau).$$

We conclude this subsection by noting, for future convenience, that we can rewrite the suppression hypothesis without referring to  $n^{0,\tau}$ :

$$\mathbb{E}_{P_{t,g}}(n_{t,g}^\tau | \tau_{t,g}^T = \rho) = \frac{1 - \xi_t \delta_{\tau \geq \rho}}{1 - \xi_t F_{t,g}^T(\tau)} \mathbb{E}_{P_{t,g}}(n_{t,g}^\tau). \tag{16}$$

### A.4 Random variables technology

Given a random variable  $X : \Omega \rightarrow \mathbb{R}$ , it is natural to consider the composition

$$\hat{X} := X \circ \sigma.$$

The main use case of this is when  $X$  represents the time at which some event related to an individual happens. For example, the infectious age  $\tau^T$  at which they get tested. In this case,  $P_t(\hat{\tau}_t^T = \tau)$  is the probability that, given an individual infected at  $t$ , their infector is tested at *the infector's* infectious age  $\tau$ .

<sup>19</sup> Note that, for simplicity, in the rest of the paper we took the distribution of  $G$  independent of absolute time, and wrote  $p_g = p_{t,g}$ .



In fact, we are often more interested in a slightly different distribution, namely that of the random variable  $\check{X}$  defined by

$$\check{X}(\omega) := X(\sigma(\omega)) - \tau^\sigma(\omega).$$

When  $X = \tau^T$ , then  $P_t(\check{\tau}_t^T = \tau)$  is now the probability that, given an individual infected at  $t$ , their source is tested at *the individual's* infectious age  $\tau$ .

The next Proposition relates the distributions of  $\hat{X}_t$ ,  $\check{X}_t$ , and  $X_t$ :

**Proposition 2** *Take  $X : \Omega \rightarrow \mathbb{R}$ . For all  $x \in \mathbb{R}$ , we have*

$$P_t(\hat{X}_t = x) = \sum_{\tau \in \mathbb{R}^+} P_{t-\tau}(X_{t-\tau} = x) \mathbb{E}_{P_{t-\tau}}(n_{t-\tau}^\tau | X_{t-\tau} = x) \frac{\#\Omega_{t-\tau}}{\#\Omega_t},$$

$$P_t(\check{X}_t = x) = \sum_{\tau \in \mathbb{R}^+} P_{t-\tau}(X_{t-\tau} = x + \tau) \mathbb{E}_{P_{t-\tau}}(n_{t-\tau}^\tau | X_{t-\tau} = x + \tau) \frac{\#\Omega_{t-\tau}}{\#\Omega_t}.$$

**Proof** We prove both formulas at once by defining  $X^\alpha : \Omega \rightarrow \mathbb{R}$  as

$$X^\alpha(\omega) := X(\sigma(\omega)) - \alpha \tau^\sigma(\omega)$$

for  $\alpha \in \{0, 1\}$ , so that  $X^0 = \hat{X}$  and  $X^1 = \check{X}$ .

$$\begin{aligned} \#\Omega_t P_t(X_t^\alpha = x) &= \#\{\omega \in \Omega_t \mid X(\sigma(\omega)) = x + \alpha \tau^\sigma(\omega)\} \\ &= \#\{(\omega', \omega) \in \tilde{\Omega} \mid \omega \in \Omega_t, X(\omega') = x + \alpha \tau^C(\omega', \omega)\} \\ &= \sum_{\tau \in \mathbb{R}^+} \#\{(\omega', \omega) \in \tilde{\Omega} \mid \omega' \in \Omega_{t-\tau}, X(\omega') = x + \alpha \tau, \omega \in \Omega_t\} \\ &= \sum_{\tau \in \mathbb{R}^+} \sum_{\omega' \in \Omega_{t-\tau} \mid X(\omega') = x + \alpha \tau} n_{t-\tau}^\tau(\omega') \\ &= \sum_{\tau \in \mathbb{R}^+} \mathbb{E}_{P_{t-\tau}}(n_{t-\tau}^\tau | X_{t-\tau} = x + \alpha \tau) \#\{\omega' \in \Omega_{t-\tau} \mid X(\omega') = x + \alpha \tau\} \\ &= \sum_{\tau \in \mathbb{R}^+} \mathbb{E}_{P_{t-\tau}}(n_{t-\tau}^\tau | X_{t-\tau} = x + \alpha \tau) P_{t-\tau}(X_{t-\tau} = x + \alpha \tau) \#\Omega_{t-\tau}. \end{aligned}$$

□

Notice that we can also break down the right hand side of the formulae of Prop. 2 by the values of  $G$ . In particular, the second equation can be rewritten as

$$P_t(\check{X}_t = x) = \sum_{\tau \in \mathbb{R}^+} \sum_{g \in \mathbb{R}} P_{t-\tau,g}(X_{t-\tau,g} = x + \alpha \tau) \mathbb{E}_{P_{t-\tau,g}}(n_{t-\tau,g}^\tau | X_{t-\tau,g})$$

$$= (x + \alpha \tau) \frac{\#\Omega_{t-\tau,g}}{\#\Omega_t}. \tag{17}$$

**A.5 Remarks on generation time and numbers of infected individuals**

In this subsection, we study the distribution of the generation time. First, we do this considering it as a function  $\tau^C : \tilde{\Omega} \rightarrow \mathbb{R}^+$  of infector–infectee pairs, and in particular taking its restriction  $\tilde{\tau}_t^C$  to the set

$$\tilde{\Omega}_t := \{(\omega', \omega) \in \tilde{\Omega} \mid \omega' \in \Omega_t\}.$$

As usual, on  $\tilde{\Omega}_t$  we put the uniform probability measure, denoted by  $\tilde{P}_t$ . In this way, we find the intuitive fact that the distribution of  $\tau^C$  restricted to  $\tilde{\Omega}_t$  is just the normalization of the infectiousness:

**Proposition 3** *The CDF of the random variable  $\tilde{\tau}_t^C : \tilde{\Omega}_t \rightarrow \mathbb{R}^+$  is given by*

$$\tilde{F}_t^C(\tau) = \frac{B_t(\tau)}{R_t}.$$

**Proof**

$$\begin{aligned} \tilde{F}_t^C(\tau) &= \tilde{P}_t(\tilde{\tau}_t^C \leq \tau) \\ &= \frac{1}{\#\tilde{\Omega}_t} \#\{(\omega', \omega) \in \tilde{\Omega}_t \mid \tau^C(\omega', \omega) \leq \tau\} \\ &= \frac{1}{\#\tilde{\Omega}_t} \sum_{\omega' \in \Omega_t} N_t^\tau(\omega') \\ &= \frac{\#\tilde{\Omega}_t}{\#\tilde{\Omega}_t} \mathbb{E}_{P_t}(N_t^\tau) \\ &= \frac{\#\tilde{\Omega}_t}{\#\tilde{\Omega}_t} B_t(\tau). \end{aligned}$$

The limit  $\tau \rightarrow +\infty$  gives

$$\#\tilde{\Omega}_t = \#\Omega_t \mathbb{E}_{P_t}(N_t^\infty) = \#\Omega_t R_t,$$

and the claim immediately follows. □

Next, we focus on the generation time as a function  $\tau_t^\sigma : \Omega_t \rightarrow \mathbb{R}^+$  of the infectee. Its probability distribution is given by the formula

$$P_t(\tau_t^\sigma = \tau) = \frac{\#\Omega_{t-\tau}}{\#\Omega_t} \mathbb{E}_{P_{t-\tau}}(n_{t-\tau}^\tau),$$

which follows from the definitions. Summing the left-hand side over all  $\tau > 0$  gives 1, from which we find how to compute  $\#\Omega_t$  in terms of quantities relative to previous times:

$$\#\Omega_t = \sum_{\tau \in \mathbb{R}^+} \#\Omega_{t-\tau} \mathbb{E}_{P_{t-\tau}}(n_{t-\tau}^\tau).$$

The last two formulae are easily proven, and the first is an immediate consequence of the next Proposition, which describes the joint probability distribution of  $\tau_t^\sigma$  and  $\hat{G}_t = (G \circ \sigma)_t$ :

**Proposition 4**

$$P_t(\tau_t^\sigma = \tau, \hat{G}_t = g) = \mathbb{E}_{P_{t-\tau,g}}(n_{t-\tau,g}^\tau) \frac{\#\Omega_{t-\tau,g}}{\#\Omega_t}.$$

This is the probability that the infector of someone infected at  $t$  was infected at  $t - \tau$  and had severity  $g$ .

**Proof**

$$\begin{aligned}
 \#\Omega_t P_t(\tau_t^\sigma = \tau, \hat{G}_t = g) &= \#\{\omega \in \Omega_t \mid \tau^\sigma(\omega) = \tau, \hat{G}(\omega) = g\} \\
 &= \#\{(\omega', \omega) \in \tilde{\Omega} \mid \omega \in \Omega_t, \omega' \in \Omega_{t-\tau}, G(\omega') = g\} \\
 &= \sum_{\omega' \in \Omega_{t-\tau, g}} n^\tau(\omega') \\
 &= \#\Omega_{t-\tau, g} \mathbb{E}_{P_{t-\tau, g}}(n_{t-\tau, g}^\tau).
 \end{aligned}$$

□

This formula for the joint probability measure will be used in the next subsection, where we will often use it in integrals. Given that in this paper we always consider  $G$  to have a given discrete range, while  $\tau^\sigma$  becomes continuous in the continuum limit, we will preferably write these integrals with respect to the improper CDFs

$$F_t^{\sigma, g}(\tau) := P_t(\tau_t^\sigma \leq \tau, \hat{G}_t = g) = \sum_{\tau' \leq \tau} \mathbb{E}_{P_{t-\tau', g}}(n_{t-\tau', g}^{\tau'}) \frac{\#\Omega_{t-\tau', g}}{\#\Omega_t}. \tag{18}$$

**A.6 Time evolution**

As we saw in Sect. 2.3, the key step to determining the time evolution of the system is writing the distribution of the notification time  $\tau_t^{A,c}$  in terms of that of the testing time  $\tau_{t'}^T$ , for  $t' < t$ . Our assumption is that any infected individual  $\omega \in \Omega_t$  is notified precisely at the testing time of their infector  $\sigma(\omega)$  with a certain probability  $s_t^c$ , provided that such testing time follows the infection time of  $\omega$  (otherwise,  $\omega$  is never notified). Referring these instants to the infectee’s infectious age, we get that  $\tau^{A,c}(\omega)$  is equal to

$$\check{\tau}^T(\omega) = \tau^T(\sigma(\omega)) - \tau^\sigma(\omega)$$

with probability  $s_t^c$  in case it is a positive number, and  $\tau^{A,c}(\omega) = +\infty$  in the remaining cases. This can be written synthetically as

$$F_t^{A,c}(\tau) = s_t^c P_t(\check{\tau}_t^T \in \mathbb{R}^+) = s_t^c (\check{F}_t^T(\tau) - \check{F}_t^T(0)), \tag{19}$$

where  $F_t^{A,c}$  and  $\check{F}_t^T$  are the improper CDFs of  $\tau_t^{A,c}$  and  $\check{\tau}_t^T$ , respectively.

Thus, our goal reduces to computing the distribution of  $\check{\tau}^T$ , the (possibly negative) time elapsed from the infectee’s contagion to the infector’s testing. Applying Eq. (17)

to  $X = \tau^T$  and using the suppression formula (16) we get, for any  $\rho \in (-\infty, +\infty]$ ,

$$\begin{aligned}
 P_t(\check{\tau}_t^T = \rho) &= \sum_{\tau \in \mathbb{R}^+} \sum_{g \in \mathbb{R}} P_{t-\tau, g}(\tau_{t-\tau, g}^T = \rho + \tau) \mathbb{E}_{P_{t-\tau, g}}(n_{t-\tau, g}^\tau | \tau^T) \\
 &= \rho + \tau \frac{\#\Omega_{t-\tau, g}}{\#\Omega_t} \\
 &= \sum_{g \in \mathbb{R}} \sum_{\tau \in \mathbb{R}^+} P_{t-\tau, g}(\tau_{t-\tau, g}^T) \\
 &= \rho + \tau \frac{1 - \xi_{t-\tau} \delta_{\tau \geq \rho + \tau}}{1 - \xi_{t-\tau} F_{t-\tau, g}^T(\tau)} \mathbb{E}_{P_{t-\tau, g}}(n_{t-\tau, g}^\tau) \frac{\#\Omega_{t-\tau, g}}{\#\Omega_t} \\
 &= \sum_{g \in \mathbb{R}} \int_{\mathbb{R}^+} P_{t-\tau, g}(\tau_{t-\tau, g}^T = \rho + \tau) \frac{1 - \xi_{t-\tau} \delta_{\rho \leq 0}}{1 - \xi_{t-\tau} F_{t-\tau, g}^T(\tau)} dF_t^{\sigma, g}(\tau).
 \end{aligned}$$

Notice that in the last line we used the improper CDFs  $F_t^{\sigma, g}$  introduced in Eq. (18).

Let us try to interpret this formula. The probability distribution of  $\check{\tau}^T$  is obtained by averaging the distributions of  $\tau_{t', g}^T$  for all  $t' = t - \tau < t$  and all  $g$ , each shifted by  $\tau$  to the left to account for the switch from the infector’s to the infectee’s infectious age. This averaging is done by integrating over all  $t'$  and  $g$  with respect to the joint distribution of the generation time  $\tau^\sigma$  and the infector’s severity  $\hat{G}$ . But a correction factor (the fraction) appears in the integral, as the fact that the infector infects at relative time  $\tau$  and has severity  $g$  conditions the distribution of  $\tau_{t-\tau, g}^T$ , by shifting it toward values greater than  $\tau$ . Indeed, the correction factor is greater than 1 for  $\rho > 0$  and less than 1 otherwise. This means that, compared to a hypothetical case in which the testing time and the infectiousness are independent (which happens when  $\xi$  is constantly zero), the probability  $P_t(\check{\tau}_t^T = \rho)$  is higher after the contagion time (i.e., when  $\rho > 0$ ) and lower before.

It follows now that the improper CDF of  $\check{\tau}_t^T$  reads

$$\check{F}_t^T(\rho) = \begin{cases} \sum_{g \in \mathbb{R}} \int_{\mathbb{R}^+} F_{t-\tau, g}^T(\rho + \tau) \frac{1 - \xi_{t-\tau}}{1 - \xi_{t-\tau} F_{t-\tau, g}^T(\tau)} dF_t^{\sigma, g}(\tau) & \text{for } \rho \leq 0, \\ \check{F}_t^T(0) + \sum_{g \in \mathbb{R}} \int_{\mathbb{R}^+} \frac{F_{t-\tau, g}^T(\rho + \tau) - F_{t-\tau, g}^T(\tau)}{1 - \xi_{t-\tau} F_{t-\tau, g}^T(\tau)} dF_t^{\sigma, g}(\tau) & \text{for } 0 < \rho < +\infty. \end{cases}$$

We only have to replace this equation in (19) to get the time evolution formula:

**Proposition 5** *Assuming the suppression hypothesis (13), the conditional independence of  $\tau^T$  and  $n_t^{0, \tau}$  with respect to  $G$ , and the notification hypothesis (19), we have*

$$F_t^{A, c}(\rho) = s_t^c \sum_{g \in \mathbb{R}} \int_{\mathbb{R}^+} \frac{F_{t-\tau, g}^T(\rho + \tau) - F_{t-\tau, g}^T(\tau)}{1 - \xi_{t-\tau} F_{t-\tau, g}^T(\tau)} dF_t^{\sigma, g}(\tau)$$

for  $\rho > 0$  and  $F_t^{A, c}(\rho) = 0$  otherwise.

### A.7 Modifications in the case of use of a contact tracing app

The inhomogeneity in the population due to the use of a contact tracing app by a part of it can be partly addressed in an analogous way to the inhomogeneity due to different degrees of severity of the illness. Namely, we introduce a new random variable

$$A : \Omega \rightarrow \{\text{app, no app}\}$$

whose value determines whether or not an individual  $\omega \in \Omega$  has the app. We assume that whether or not an individual has the app is independent of both their severity and their infectiousness *in the absence* of measures. In other words,  $A$  is independent of  $G$  and  $n^{0,\tau}$ , for all  $\tau$ . On the other hand, the infectiousness (in presence of measures) and the testing time of an individual will be different depending on whether or not they use the app.

A further partitions  $\Omega$ : we write

$$\Omega_{t,g,a} := \{\omega \in \Omega_t \mid G(\omega) = g, A(\omega) = a\}, \quad n_{t,g,a}^\tau := n^\tau \upharpoonright_{\Omega_{t,g,a}},$$

and so on. The content of Sect. A.3 fully applies to this scenario, but we want now to have formulae conditioned on  $A$ . As  $A$  and  $n^{0,\tau}$  are independent, the previous formulae simply become

$$\mathbb{E}_{P_{t,g,a}}(n_{t,g,a}^\tau) = \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau})(1 - \xi_t F_{t,g,a}^T(\tau))$$

and

$$\begin{aligned} \mathbb{E}_{P_{t,g,a}}(n_{t,g,a}^\tau \mid \tau_{t,g,a}^T = \rho) &= (1 - \xi_t \delta_{\tau \geq \rho}) \mathbb{E}_{P_{t,g}}(n_{t,g}^{0,\tau}) \\ &= \frac{1 - \xi_t \delta_{\tau \geq \rho}}{1 - \xi_t F_{t,g,a}^T(\tau)} \mathbb{E}_{P_{t,g,a}}(n_{t,g,a}^\tau). \end{aligned} \tag{20}$$

The suppression formula for  $R_t$  can be broken down to

$$\begin{aligned} R_t &= \sum_{g \in \mathbb{R}} \sum_a P_{t,g} \epsilon_{t,a} R_{t,g,a} \\ &= \sum_g \sum_a P_{t,g} \epsilon_{t,a} \int_{\mathbb{R}^+} (1 - \xi_t F_{t,g,a}^T(\tau)) dB_{t,g}^0(\tau), \end{aligned}$$

where  $\epsilon_{t,a} := P_t(A_t = a)$ .

The time evolution equation has to be treated differently, as the receipt of the notification depends on whether both the infector and the infectee use the app.

Let  $\check{F}_t^{T,a}(\tau)$  denote the probability that, given an individual  $\omega \in \Omega_t$ , the infector  $\sigma(\omega)$  is tested at a time  $\leq t + \tau$  and we have  $\hat{A}(\omega) = A(\sigma(\omega)) = a$ .

According to our assumptions, given an infection occurred at  $t$ , the probability that the infector notifies the infectee when they test positive (provided that this happens after the infection) is  $s_t^{c,\text{app}}$  in the case that both individuals have the app, and is  $s_t^{c,\text{no app}}$  otherwise. Therefore, the contact tracing hypothesis (19) is now replaced by the following expressions for the CDFs of the time of the notification received by an

individual with or without the app, respectively:

$$\begin{aligned}
 F_{t,\text{app}}^{A,c}(\rho) &= s_t^{c,\text{app}}(\check{F}_t^{\text{T,app}}(\rho) - \check{F}_t^{\text{T,app}}(0)) + s_t^{c,\text{no app}}(\check{F}_t^{\text{T,no app}}(\rho) - \check{F}_t^{\text{T,no app}}(0)), \\
 F_{t,\text{no app}}^{A,c}(\rho) &= s_t^{c,\text{no app}}(\check{F}_t^{\text{T,app}}(\rho) - \check{F}_t^{\text{T,app}}(0) + \check{F}_t^{\text{T,no app}}(\rho) - \check{F}_t^{\text{T,no app}}(0)) \\
 &= s_t^{c,\text{no app}}(\check{F}_t^{\text{T}}(\rho) - \check{F}_t^{\text{T}}(0)).
 \end{aligned}
 \tag{21}$$

Now, the improper CDFs  $\check{F}_t^{\text{T},a}$  can be computed just as before, simply treating the conditioning on  $\hat{A}$  as we treated the conditioning on  $\hat{G}$ :

$$\begin{aligned}
 P_t(\check{\tau}_t^{\text{T}} = \rho, \hat{G}_t = g, \hat{A}_t = a) \\
 = \sum_{\tau > 0} P_{t-\tau,g,a}(\tau_{t-\tau,g,a}^{\text{T}} = \rho + \tau) \mathbb{E}_{P_{t-\tau,g,a}}(n_{t-\tau,g,a}^{\tau} | \tau^{\text{T}} = \rho + \tau) \frac{\#\Omega_{t-\tau,g,a}^{\rho}}{\#\Omega_t^{\rho}} \tag{22} \\
 = \int_{\mathbb{R}^+} P_{t-\tau,g,a}(\tau_{t-\tau,g,a}^{\text{T}} = \rho + \tau) \mathbb{E}_{P_{t-\tau,g,a}}(n_{t-\tau,g,a}^{\tau} | \tau^{\text{T}} = \rho + \tau) dF_t^{\sigma,g,a}(\tau),
 \end{aligned}$$

where we defined  $F_t^{\sigma,g,a}$  as follows, proceeding like in Sect. A.5 to compute the joint distribution of  $\tau_t^{\sigma}$ ,  $\hat{G}_t$ , and  $\hat{A}_t$ :

$$F_t^{\sigma,g,a}(\tau) := P_t(\tau_t^{\sigma} \leq \tau, \hat{G}_t = g, \hat{A}_t = a) = \sum_{\tau' \leq \tau} \mathbb{E}_{P_{t-\tau',g,a}}(n_{t-\tau',g,a}^{\tau'}) \frac{\#\Omega_{t-\tau',g,a}^{\tau}}{\#\Omega_t^{\tau}}. \tag{23}$$

It is worth noting that comparing this equation with (18) we get

$$\frac{dF_t^{\sigma,g,a}(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g,a}^{\text{T}}} = \epsilon_{t,a} \frac{dF_t^{\sigma,g}(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g}^{\text{T}}}. \tag{24}$$

Replacing the suppression formula (20) in (22) and summing over  $\rho$ , we end up with

$$\check{F}_t^{\text{T},a}(\rho) - \check{F}_t^{\text{T},a}(0) = \sum_g \int_{\mathbb{R}^+} \frac{F_{t-\tau,g,a}^{\text{T}}(\rho + \tau) - F_{t-\tau,g,a}^{\text{T}}(\tau)}{1 - \xi_{t-\tau} F_{t-\tau,g,a}^{\text{T}}} dF_t^{\sigma,g,a}(\tau)$$

for  $\rho > 0$ . Plugging this into (21) gives us  $F_{t,a}^{A,c}$  in terms of  $F_t^{\sigma,g,a}$  and  $F_{t',g,a}^{\text{T}}$  for  $t' < t$ , that is the time evolution equation for the scenario with app usage. For  $a = \text{app}$ , this is Eq. (9), while for  $a = \text{no app}$  it simplifies to Eq. (10), since when the infectee doesn't have the app it is irrelevant whether or not the infector has the app. This is evident from the last line of Eq. (21), which in fact could also have been used, together with the expression for  $\check{F}_t^{\text{T}}$  derived in Sect. A.6, to get Eq. (10). Using Eq. (24), it can be checked immediately that the two approaches give the same result.

### References

Bendavid E, Mulaney B, Sood N, Shah S, Bromley-Dulfano R, Lai C, Weissberg Z, Saavedra-Walker R, Tedrow J, Bogan A, Kupiec T, Eichner D, Gupta R, Ioannidis JPA, Bhattacharya J (2021) COVID-19

- antibody seroprevalence in Santa Clara County, California. *Int J Epidemiol*. <https://doi.org/10.1093/ije/dyab010>
- Cori A, Ferguson NM, Fraser C, Cauchemez S (2013) A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* 178(9):1505–1512
- 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*. <https://doi.org/10.1126/science.abb6936>
- Fraser C, Riley S, Anderson RM, Ferguson NM (2004) Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci* 101(16):6146–6151. <https://doi.org/10.1073/pnas.0307506101>
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. <https://doi.org/10.1056/NEJMoa2001316>
- Maiorana A, Meneghelli M (2021) Epidemics suppression model. <https://github.com/BendingSpoons/epidemic-suppression-model>
- Mizumoto K, Kagaya K, Zarebski A, Chowell G (2020) Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* 25(10):2000180. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180>
- Müller J, Kretzschmar M, Dietz K (2000) Contact tracing in stochastic and deterministic epidemic models. *Math Biosci* 164(1):39–64. [https://doi.org/10.1016/S0025-5564\(99\)00061-9](https://doi.org/10.1016/S0025-5564(99)00061-9)
- Pathogen Dynamics Group OU Big Data Institute (2020) <https://github.com/BDI-pathogens/OpenABM-Covid19>
- 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. <https://doi.org/10.1098/rsos.202091>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.