
This is the **accepted version** of the journal article:

Barril, Carles; Calsina i Ballesta, Àngel; Cuadrado Gavilán, Sílvia; [et al.].
«Reproduction number for an age of infection structured model». Mathematical Modelling of Natural Phenomena, Vol. 16 (June 2021), art. 42. DOI 10.1051/mmnp/2021033

This version is available at <https://ddd.uab.cat/record/268832>

under the terms of the  license

REPRODUCTION NUMBER FOR AN AGE OF INFECTION STRUCTURED MODEL *

CARLES BARRIL¹, ÀNGEL CALSINA^{1,2}, SÍLVIA CUADRADO¹ AND JORDI RIPOLL³

Abstract. We study the basic reproduction number (R_0) in an epidemic model where infected individuals are initially asymptomatic and structured by the time since infection. At the beginning of an epidemic outbreak the computation of R_0 relies on limited data based mostly on symptomatic cases, since asymptomatic infected individuals are not detected by the surveillance system. R_0 has been widely used as an indicator to assess the dissemination of infectious diseases. Asymptomatic individuals are assumed to either become symptomatic after a fixed period of time or they are removed (recovery or disease-related death). We determine R_0 understood as the expected secondary symptomatic cases produced by a symptomatic primary case through a chain of asymptomatic infections. R_0 is computed directly by interpreting the model ingredients and also using a more systematic approach based on the next-generation operator. Reported Covid-19 cases data during the first wave of the pandemic in Spain are used to fit the model and obtain both values of R_0 before and after the severe lockdown imposed in March 2020. The results confirm that SARS-CoV-2 was expanding within the population before the lockdown whereas the virus spreading was controlled two weeks after the lockdown.

In memoriam Carles Perelló,
emeritus professor at UAB, recently deceased and scientific father/grandfather of the authors.

Mathematics Subject Classification. 35Q92, 92D30.

The dates will be set by the publisher.

CONTENTS

1.	Introduction	2
2.	Exponential Growth	4
3.	Basic reproduction number	6
4.	Next-Generation Operator	7
5.	Simplifications and Covid-19	9
6.	Discussion	11

Keywords and phrases. age of infection; basic reproduction number; covid-19

* This research was funded by Ministerio de Ciencia e Innovación, grants number MTM2017-84214-C2-2-P and RED2018-102650-T

¹ Departament de Matemàtiques, Universitat Autònoma de Barcelona; e-mail: carlesbarril@mat.uab.cat, acalsina@mat.uab.cat, silvia@mat.uab.cat

² Centre de Recerca Matemàtica

³ Departament d'Informàtica, Matemàtica Aplicada i Estadística, Universitat de Girona; e-mail: jripoll@imae.udg.edu

© EDP Sciences, 2021

1. INTRODUCTION

Emergence of infectious diseases is a problem that humans are doomed to face over and over. Beyond the particularities of each new pathogen, the dynamics of epidemics depend on similar processes. This makes mathematical models a useful tool to evaluate the possible outcomes as a function of certain key parameters.

The expansion of an epidemics is usually monitored by means of two indicators: the exponential growth rate of the disease and the basic reproduction number, denoted by R_0 . The former corresponds to the change in the number of active cases normalised by the prevalence at a given time, so that negative exponential growth rates indicate that the number of infected individuals is decaying exponentially whereas positive exponential growth rates mean that the number of infections is growing exponentially. The latter indicator, R_0 , is usually defined as the number of new infections an infected individual will produce in a wholly susceptible population, so that if it is larger than one each infected individual produces more than one secondary infection and the disease is able to disseminate whereas if it is smaller than one this cannot happen (see [11] for an historical review on R_0). As expected by the above definitions, the sign of the exponential growth rate coincides with the sign of $R_0 - 1$ (in [20] a proof of this relation is proven for deterministic models dealing with infinite dimensional variables, as the model analysed in this article). Although the two indicators seem to be equivalent in regard to the analysis of the epidemic, the mechanistic interpretation of R_0 and its, often, explicit dependence on relevant parameters of the model, makes it especially well suited to design and evaluate social measures intended to control and eradicate the epidemics [9].

As it has been discussed elsewhere different reproduction numbers can be defined depending on what is understood as a birth event, which in an epidemiological context could correspond to a new infection or, alternatively, to a new reported case [5, 9]. The choice of R_0 could be motivated by the strategy used to control the epidemics [9]. For example, at the beginning of a pandemic only reported cases can be used to monitor how it spreads, so that it is natural to consider to focus on how the number of reported cases grow.

In this study we consider an extension of the infection-age structured compartmental model introduced by Kermack and McKendrick in 1927 ([15]) where we take into account the existence of individuals that experience the disease with no (or very mild) symptoms making them undetectable for the health system as long as systematic testing is not being applied to the population.

That is, we divide the infected population into two groups: asymptomatic individuals (we also include here symptomatic with mild symptoms) for which no isolation measures have been taken and symptomatic individuals. The role of infected asymptomatic individuals seems to be very important for certain diseases [7, 13] and in particular for Covid-19 (see for instance [2, 8, 9, 13] for models considering asymptomatic individuals and [1, 16] for applications to Covid-19).

We denote by $S(t)$ the number of individuals that are susceptible to the disease. Let τ denote the age of infection (time elapsed since infection) and $i(\tau, t)$ the density with respect to τ of infected individuals at time t that are asymptomatic (here the term “asymptomatic” refers to individuals without symptoms at a given time, which could or could not develop symptoms in the future). We assume that a newly infected individual is asymptomatic. $J(t)$ stands for the number of symptomatic individuals (detected by the health system). Finally $R(t)$ denotes the number of individuals who have been infected and then removed from the chain of infection (in the sense that they cannot be infected again). Thus, $R(t)$ are either asymptomatic individuals that recovered without being detected or symptomatic individuals that recovered or symptomatic individuals that died due to the disease.

The usual mass-action incidence is assumed. Specifically, denoting by $\beta_1(\tau)$ the transmission rate of asymptomatic individuals (i.e. the number of contacts per unit of time of an asymptomatic individual multiplied by the probability that the contact between an asymptomatic individual and a susceptible one results in transmission of the disease) and by N the total population (see (2) below) we have that $\frac{S(t)}{N}$ is the probability that

the contact is indeed with a susceptible¹ and therefore $\beta_1(\tau)\frac{S(t)}{N}$ gives the number of susceptible individuals that become infected per unit of time per each infectious asymptomatic individual of infection-age τ . Notice in particular that the model includes the possibility of latency (or incubation time) by assuming $\beta_1(\tau)$ vanishing for small infection-age. Analogously, denoting by β_2 the transmission rate of symptomatic individuals, $\beta_2\frac{S(t)}{N}$ will be the number of susceptible individuals that become infected per unit of time per symptomatic individual. Whereas some other parameters of the model can usually be measured experimentally, both $\beta_1(\tau)$ and β_2 are difficult to measure directly because they depend on the number of contacts and the probability of transmission, which are not easy to determine. Since symptomatic cases are aware of being infectious, they presumably will avoid contacts with others (i.e. they will self-isolate) making β_2 to be smaller than β_1 (this is observed in section 5 where real data is analysed). The recovery rate of an asymptomatic individual of age τ is given by $\gamma_1(\tau)$ and γ_2 is the so-called recovery rate of the symptomatic population (which is in fact the sum of the actual recovery rate and the disease-related death rate).

The model reads:

$$\begin{aligned} S'(t) &= -S(t) \int_0^T \frac{\beta_1(\tau)}{N} i(\tau, t) d\tau - S(t) \frac{\beta_2}{N} J(t), \\ i_t(\tau, t) + i_\tau(\tau, t) + \gamma_1(\tau) i(\tau, t) &= 0, \quad \tau \in [0, T], \\ J'(t) &= p i(T, t) - \gamma_2 J(t), \\ R'(t) &= \int_0^T \gamma_1(\tau) i(\tau, t) d\tau + (1 - p) i(T, t) + \gamma_2 J(t), \\ i(0, t) &= S(t) \int_0^T \frac{\beta_1(\tau)}{N} i(\tau, t) d\tau + S(t) \frac{\beta_2}{N} J(t), \end{aligned} \tag{1}$$

where T is the time after which asymptomatic individuals which have not recovered yet, either, with probability p develop symptoms and are detected by the health system passing then to the J -compartment or, with probability $1 - p$ recover and become immune passing then to the R -compartment. The sudden onset of severe symptoms seems to be a characteristic feature of the Covid-19 [19]. The subscripts t and τ in the second equation denote the partial derivative of function i with respect t and τ respectively. All parameters are non negative and β_2 , γ_2 and p are assumed to be strictly positive to avoid some limiting cases.

Notice that we have assumed that the time scale of the epidemic is much faster than the time scale for demographic processes (natural birth and death), and therefore these are not included in the model and the total population is then constant and given by

$$N = S(t) + \int_0^T i(\tau, t) d\tau + J(t) + R(t) \quad \text{for all } t \geq 0. \tag{2}$$

Notice also that the dynamics of S , i and J are independent of the variable R . Hence in what follows we will concentrate on the three first equations of (1) plus the boundary condition (the fifth equation), noting that from the knowledge of $S(t)$, $i(t)$ and $J(t)$ one easily can compute $R(t)$ integrating the fourth equation in time or directly using that the total population is constant, i.e. equation (2).

The paper is structured as follows: in Section 2 we compute the characteristic equation of the linearization of the model around the disease free equilibrium and give a threshold for the spectral bound of the linear operator and so a threshold for the stability/instability of this steady state. In Section 3 we compute the so-called basic reproduction number $R_{0,s}$ defined here as the expected number of symptomatic individuals produced by a single newly symptomatic one and check the well known general result that $R_{0,s}$ is larger than one if and only if the spectral bound is positive, and less than one if and only if the spectral bound is negative. In Section 4 we use results on the reproduction number in continuously structured populations to compute $R_{0,s}$ as the

¹Actually this is an approximation since the total population N includes a small fraction of deceased individuals by the disease.

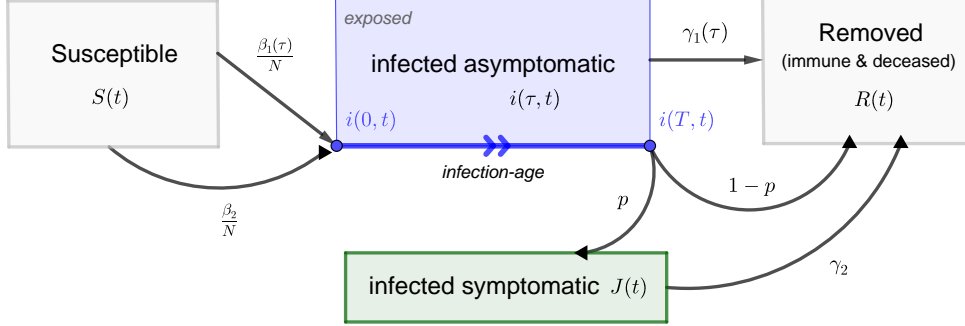


FIGURE 1. Flow diagram representing the dynamics of the model (1). Per capita new (hidden) infections per unit of time (arrows departing from the Susceptible class) are $\int_0^T \frac{\beta_1(\tau)}{N} i(\tau, t) d\tau$ and $\frac{\beta_2}{N} J(t)$, produced by asymptomatic and symptomatic individuals respectively, whereas new (physically observable) infections per unit of time are $p i(T, t)$. Exposed individuals can be included into the asymptomatic class for small ages where $\beta_1(\tau)$ is zero.

spectral bound of the so-called next generation operator. Finally, in Section 5 we consider a simplified model, determine parameters based on the Spanish record data of the Covid-19 pandemic during the springtime peak and compute the corresponding values of $R_{0,s}$ before and immediately after the lockdown. The last section is devoted to some concluding remarks.

2. EXPONENTIAL GROWTH

Let us consider the (initial) situation when the number of infected individuals is very small compared to the (possibly fully) susceptible population. Then the dynamics of the infected population is described by the linearized equations for the infected population at the disease-free steady state $(N, 0, 0, 0)$ of (1) which are

$$\begin{aligned} i_t(\tau, t) + i_\tau(\tau, t) + \gamma_1(\tau) i(\tau, t) &= 0, \quad \tau \in [0, T], \\ i(0, t) &= \int_0^T \beta_1(\tau) i(\tau, t) d\tau + \beta_2 J(t), \\ J'(t) &= p i(T, t) - \gamma_2 J(t). \end{aligned} \tag{3}$$

System (3) can be written as

$$\frac{d}{dt} z(t) = Bz(t) - Mz(t),$$

with

$$z(t)(\tau) := \begin{pmatrix} i(\tau, t) \\ J(t) \end{pmatrix}, \quad B := \begin{pmatrix} 0 & 0 \\ p\delta_T & 0 \end{pmatrix}, \quad M := \begin{pmatrix} \partial_\tau + \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}, \tag{4}$$

where δ_T stands for the evaluation operator at the point $\tau = T$ and the first order differential operator M is supplemented with the boundary condition given by the second equation in (3). The main reason to consider the decomposition in operators B and M is that it allows to compute the so-called next generation operator when the “birth event” is the emergence of a symptomatic case (see section 4).

Since $-M$ is the infinitesimal generator of a positive semigroup in $L^1(0, T) \times \mathbb{R}$ with domain

$$D_M = \left\{ (i, J) \in W^{1,1}(0, T) \times \mathbb{R} : i(0) = \int_0^T \beta_1(\tau) i(\tau) d\tau + \beta_2 J \right\} \quad (5)$$

and B is a bounded linear operator in the domain of M endowed with the graph norm we get that $B - M$ is the infinitesimal generator of a positive semigroup (see Section 3.2 in [5]).

Since the equations in (3) are linear we look for separable solutions, i.e. solutions of the form

$$\begin{pmatrix} i(\tau, t) \\ J(t) \end{pmatrix} = e^{\lambda t} \begin{pmatrix} y(\tau) \\ j \end{pmatrix}.$$

Substituting them in (3) we obtain

$$\begin{aligned} \lambda y(\tau) + y'(\tau) + \gamma_1(\tau) y(\tau) &= 0, \\ y(0) &= \int_0^T \beta_1(\tau) y(\tau) d\tau + \beta_2 j, \\ \lambda j &= p y(T) - \gamma_2 j. \end{aligned} \quad (6)$$

The solution of the first equation in (6) is $y(\tau) = y(0)e^{-\lambda\tau - \Gamma(\tau)}$, where $\Gamma(\tau) = \int_0^\tau \gamma_1(s) ds$. Substituting it in the second equation in (6) yields

$$y(0) = y(0) \int_0^T \beta_1(\tau) e^{-\lambda\tau - \Gamma(\tau)} d\tau + \beta_2 j, \quad (7)$$

and substituted in the third equation in (6) gives

$$\lambda j = p y(0) e^{-\lambda T - \Gamma(T)} - \gamma_2 j. \quad (8)$$

Equations (7) and (8) constitute a linear homogeneous system for the unknown $(y(0), j)^t$ with nontrivial solution if and only if λ is a solution (real or complex) of the characteristic equation

$$p e^{-\lambda T - \Gamma(T)} \beta_2 - (\gamma_2 + \lambda) \left(1 - \int_0^T \beta_1(\tau) e^{-\lambda\tau - \Gamma(\tau)} d\tau \right) = 0,$$

or, equivalently, if λ is a solution of

$$f(\lambda) := \frac{p e^{-\lambda T - \Gamma(T)} \beta_2}{(\gamma_2 + \lambda) \left(1 - \int_0^T \beta_1(\tau) e^{-\lambda\tau - \Gamma(\tau)} d\tau \right)} = 1. \quad (9)$$

Since the solutions of (3) with positive initial conditions are positive ($B - M$ generates a positive semigroup), the maximum of the real part of the roots of the characteristic equation is a (real) root of it (the spectral bound of a positive semigroup is a spectral value). Moreover, as the exponential behaviour of the semigroup generated by $B - M$ is determined by the spectral bound of $B - M$, the larger real solution of the characteristic equation determines the exponential growth or decay of the solution [12].

Notice that the second factor in the denominator of $f(\lambda)$ is an increasing function with limit 1 when λ tends to ∞ . Moreover f is a strictly decreasing continuous function with limit 0 when λ tends to ∞ whenever this second factor is positive and $\lambda > -\gamma_2$.

Let us consider two cases:

- $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau \geq 1$.

In this case f has a positive singularity (or equal to zero) and for λ bigger than this singularity, f decreases from ∞ to zero. Therefore there exists at least one positive value of λ such that $f(\lambda) = 1$. This implies the existence of solutions that grow exponentially, i.e., the disease can invade the susceptible population. Notice that this is independent in this case of the values of the parameters of the symptomatic population. On the other hand, a look at equation (9) confirms that its real solution λ can be made arbitrarily large by increasing the product $p\beta_2$.

Since from (8) we have that the proportion of symptomatic individuals to the asymptomatic ones is given by

$$\frac{j}{y(0)} = \frac{pe^{-\lambda T - \Gamma(T)}}{\lambda T + \gamma_2},$$

it follows that large values of β_1 (to have the integral larger than 1) and of β_2 (to have the solution of (9) large) could lead to a quick spreading of the epidemics with, initially, very small numbers of symptomatic individuals compared to the asymptomatic ones (see the Discussion).

- $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau < 1$.

In this case f is a continuous function on $[0, \infty)$, decreasing for $\lambda > 0$ and with limit 0 when λ tends to ∞ which implies that (9) has a positive solution if and only if $f(0) > 1$. That means that there will be solutions growing exponentially if $f(0) > 1$ whereas if $f(0) < 1$ all solutions will decrease exponentially.

Summarizing the discussion above we have:

Let

$$f(0) = \frac{\beta_2}{\gamma_2} \frac{1}{1 - \int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau} e^{-\Gamma(T)} p. \quad (10)$$

- If $f(0) < 0$ or $f(0) > 1$, the infected population grows exponentially.
- If $f(0) \in (0, 1)$, the infected population decreases exponentially.

3. BASIC REPRODUCTION NUMBER

In this section we show that the threshold parameter that we have defined in the previous section (10) is indeed the expected number of symptomatic individuals that a symptomatic individual will produce provided that $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau < 1$.

First notice that $\frac{1}{\gamma_2}$ is the expected duration of the symptomatic period and β_2 is the number of infections per unit of time that a symptomatic individual produces.

Thus the expected number of susceptible individuals infected by a symptomatic individual and becoming then asymptomatic infected individuals is

$$\frac{\beta_2}{\gamma_2}. \quad (11)$$

Each one of them will remain infected without symptoms for a time bigger than $\tau (\leq T)$ with probability $e^{-\Gamma(\tau)}$ and, during the interval $(\tau, \tau + d\tau)$, will infect $\beta_1(\tau)d\tau$ susceptible individuals, in such a way that, on average, $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau$ new asymptomatic individuals will be produced. Each one of them will again infect, on average, the same number of susceptibles, and the process will continue in a recursive way.

Therefore, the expected total number of infected asymptomatic individuals produced by each secondary asymptomatic infected by the original symptomatic individual will be, counting the first asymptomatic as well

and assuming that $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau < 1$,

$$\sum_{k=0}^{\infty} \left(\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau \right)^k = \frac{1}{1 - \int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau}, \quad (12)$$

whereas they will be infinitely many in the opposite case.

For each one of these individuals the probability of reaching the end of the asymptomatic period and becoming then symptomatic is

$$e^{-\Gamma(T)}p. \quad (13)$$

The product of (11), (12) and (13) gives the following (see [9] for the same argument in a simpler model):

$$R_{0,s} = \frac{\beta_2}{\gamma_2} \frac{1}{1 - \int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau} e^{-\Gamma(T)}p \quad (14)$$

whenever $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau < 1$, so that the right hand side of (10) coincides with $R_{0,s}$.

Notice that in the case that $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau \geq 1$, i.e. when the average number of secondary infections produced by an infected individual during the asymptomatic phase is greater than one, the common ratio of the previous geometric sequence is bigger than or equal to 1 and therefore the expected number of infected individuals produced by each asymptomatic individual would be infinite (not by direct contact but counting all “generations”). The basic reproduction number would in this case be infinite (as in [9] and [4]). However we should not forget on the one hand that we are not taking into account how much time it takes so that all the infections happen (which could be very big) and on the other hand that we are analysing a linearised model in which we have assumed that the number of susceptibles remains constant. When this is not the situation anymore, the nonlinear model (1) has to be analysed with different mathematical tools.

In addition, analogously to the derivation above, we can compute the asymptomatic reproduction number (the expected secondary asymptomatic cases produced by an asymptomatic primary case) as:

$$R_{0,a} = \int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau + \frac{\beta_2 p}{\gamma_2} e^{-\Gamma(T)}$$

since the first term corresponds to the infections produced by the primary case during the asymptomatic phase and the second term corresponds to the infections produced by the primary case during the symptomatic phase (provided it becomes symptomatic) multiplied by the probability that it becomes symptomatic. This is the reproduction number commonly used in epidemiology and defined as the expected number of new infections a newly infected individual will produce in a wholly susceptible population over the full course of the disease. Notice that $R_{0,a} > 1$ if and only if $R_{0,s} > 1$ and $R_{0,a} < 1$ if and only if $R_{0,s} < 1$.

4. NEXT-GENERATION OPERATOR

Another way, more systematic, to obtain the basic reproduction number R_0 consists in computing it as the spectral radius of the so-called *next-generation operator* ([3, 10, 14]). For the numerical computation of R_0 see the efficient method introduced in [6]. The next generation operator is defined as BM^{-1} where M (transition/mortality operator) is a linear operator such that $-M$ is the generator of a strongly continuous semigroup of positive linear operators, whose spectral bound is strictly negative ($s(-M) < 0$) and B (birth/infections operator) is a positive linear operator such that $B - M$ (the operator defining the original model considered) is the infinitesimal generator of a positive semigroup.

Such a decomposition of the generator is not generally unique and therefore one can define different next-generation operators and consequently different basic reproduction numbers for the same model. Nevertheless

for all of them one has that the solutions of the linear system will grow exponentially if $R_0 > 1$ and decrease exponentially if $R_0 < 1$ ([9, 10, 20]).

For system (3) we have defined the operators B and M in (4). Assuming that $s(-M) < 0$ and so the operator M is invertible, since BM^{-1} is a one dimensional range operator we will have that $R_{0,s}$ coincides with the second component of

$$BM^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

To find the preimage of $(0, 1) \in L^1(0, T) \times \mathbb{R}$ under the linear operator M with domain given in (5), we consider

$$\begin{pmatrix} i(\tau) \\ J \end{pmatrix} \in D_M \quad \text{such that} \quad \begin{pmatrix} i(\tau) \\ J \end{pmatrix} = M^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

which implies, by applying M to both sides,

$$\begin{pmatrix} i'(\tau) + \gamma_1(\tau)i(\tau) \\ \gamma_2 J \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow \begin{matrix} i(\tau) = i(0)e^{-\int_0^\tau \gamma_1(\sigma)d\sigma} \\ J = 1/\gamma_2 \end{matrix}.$$

Since $i(0) = \int_0^T \beta_1(\tau)i(\tau)d\tau + \beta_2 J$ because the pair $(i, J) \in D_M$, the value of $i(0)$ can be expressed in terms of the parameters:

$$i(0) = \frac{\beta_2}{\gamma_2} \frac{1}{1 - \int_0^T \beta_1(s)e^{-\Gamma(s)}ds}.$$

Therefore,

$$M^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\beta_2}{\gamma_2} \frac{1}{1 - \int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau} e^{-\Gamma(\cdot)} \\ \frac{1}{\gamma_2} \end{pmatrix},$$

and

$$R_{0,s} = \frac{\beta_2}{\gamma_2} \frac{1}{\left(1 - \int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau\right)} e^{-\Gamma(T)} p. \quad (15)$$

The condition on the spectral bound of $-M$ coincides with the condition $\int_0^T \beta_1(\tau)e^{-\Gamma(\tau)}d\tau < 1$ above.

Just to show the flexibility of the introduced age-dependent model (1) in describing different epidemic scenarios, let us illustrate the following biologically meaningful example. We can assume an increasing infection transmission, e.g. $\beta_1(\tau) = \bar{\beta} \cdot (\tau - l)$, $\bar{\beta} > 0$, with support $[l, T]$, since the viral load in the host eventually increases in time until the appearance of the first symptoms. On the other hand, for the recovery and disease-induced mortality, we can take one of the survival probabilities mostly used in population dynamics, e.g. either $\Gamma(\tau) = e^{-\bar{\gamma}_1 \tau}$, $\bar{\gamma}_1 > 0$, or $\Gamma(\tau) = (1 - \tau/\bar{T})^\alpha$, $\bar{T} \geq T$, $\alpha > 0$. For instance, taking the former Γ and β_1 with $l = 0$ above, we get the basic reproduction number from (15) as

$$R_{0,s} = \frac{\beta_2 p}{\gamma_2} \frac{\bar{\gamma}_1^2 e^{-\bar{\gamma}_1 T}}{\bar{\gamma}_1^2 - \bar{\beta} [1 - (1 + \bar{\gamma}_1 T)e^{-\bar{\gamma}_1 T}]},$$

which is well-defined whenever $\bar{\beta} \leq \bar{\gamma}_1^2$ and it is a decreasing function of T , ranging from $\frac{\beta_2 p}{\gamma_2}$, when there is no asymptomatic period ($T = 0$), to vanish when the asymptomatic period is very large.

Interestingly enough, the behaviour of $R_{0,s}$ as a function of the asymptomatic period T is very different when $\bar{\beta} > \bar{\gamma}_1^2$. In this case it is still decreasing for small T (until $T = \frac{1}{\bar{\gamma}_1} \ln(\frac{\bar{\beta}}{\bar{\beta} - \bar{\gamma}_1^2})$) but for larger values of T it increases and becomes unbounded for a finite value of T . Summarizing, a longer asymptomatic period reduces the basic reproduction number (and so it is beneficial) when the infectiousness of asymptomatic individuals is small compared to their rate of recovery but in the contrary case, it can become very harmful, to the extreme of making the pandemic uncontrollable with actions directed only at the population of symptomatic individuals.

5. SIMPLIFICATIONS AND COVID-19

The main practical difficulty of models like (1) is not really the mathematical analysis that, at least, could be done numerically, but the estimation of the parameters involved, especially functional parameters like β_1 and γ_1 in (1).

In order to simplify the situation we can assume that β_1 is constant and that $\gamma_1 = 0$. This means that asymptomatic disease transmission does not change during the period without symptoms and that asymptomatic individuals do not recover before a fixed time T .

More precisely, β_1 constant means that we are assuming that the number of contacts of asymptomatic individuals does not change, which could be a natural hypothesis when neglecting environmental changes. It also means though that the viral load (and therefore the success of infection in a contact) does not change which is much more arguable. However we assume it here for the sake of simplicity.

Under these assumptions, from the results in section 2 we have that $\beta_1 T \geq 1$ implies exponential growth of the disease and $R_{0,s} = \infty$ whereas if $\beta_1 T < 1$ we have from (14),

$$R_{0,s} := \frac{\beta_2 p}{\gamma_2(1 - \beta_1 T)}. \quad (16)$$

Notice that, if $p = 1$ and $T = 0$, i.e. if the asymptomatic period is neglected, $R_{0,s}$ reduces to the basic reproduction number of the ODE Kermack-McKendrick SIR model).

We can distinguish different kinds of parameters in (16). p , T and γ_2 are intrinsic parameters of the disease: p is the probability that an infected asymptomatic individual becomes symptomatic, T is the time from infection until the development of symptoms (or until recovery) and γ_2^{-1} is the average length of the symptomatic period. Since SARS-CoV-2 is a new virus, all of them are pretty unknown. However, they are probably rather fixed until a treatment is found and in principle they have similar values in all countries affected by the disease.

On the other hand β_1 and β_2 are epidemiological parameters that depend on the number of contacts, which might differ in each country because of socio-economic and cultural reasons. Moreover these are not constant parameters, they can be, and have been controlled with quarantine measures.

From (3) we see that i is a hidden variable and that the available daily data corresponds to the new “detected” cases and to the sum of disease-related deaths and recovered, which amount (approximately) to $p i(T, t)$ and $\gamma_2 J(t)$ respectively. We denote the new detected cases by $D(t)$, i.e. $D(t) := p i(T, t)$. The knowledge of p and T would allow to find $i(0, t)$ (daily new infected cases) as well as β_1 and β_2 .

Under the assumption that γ_1 vanishes identically, the first equation in (3) can be trivially integrated yielding $i(\tau, t) = i(0, t - \tau)$. So, using the second equation in (3) we can write

$$\begin{aligned} D(t) &= p i(T, t) = p i(0, t - T) = \beta_1 \int_0^T p i(\tau, t - T) d\tau + p \beta_2 J(t - T) \\ &= \beta_1 \int_0^T p i(T, t - \tau) d\tau + p \beta_2 J(t - T) = \beta_1 \int_{t-T}^t D(s) ds + \beta_2 p J(t - T). \end{aligned} \quad (17)$$

A procedure to estimate the parameters β_1 and β_2 would proceed as follows. Let us assume we have a daily series of values of new cases $D_i := D(t_i)$ and also the total number of symptomatic individuals $J_i := J(t_i)$ beginning at time $-T + 1 =: t_{-T+1}$ and ending at some positive time $t = t_N$. The trapezoidal rule to approximate the integral in (17) for $t \in [0, t_N]$ gives, for $j = 1, \dots, N$, and assuming that T (an integer number of days) and p are known,

$$D_j = \beta_1 \left(\frac{1}{2} D_{j-T} + \sum_{i=1}^{T-1} D_{j-T+i} + \frac{1}{2} D_j \right) + \beta_2 p J_{j-T} =: \beta_1 X_j + \beta_2 Y_j, \quad (18)$$

which is an overdetermined linear system (if $N > 2$) for β_1 and β_2 . The method of least squares gives the following formulas for these parameters:

		daily new cases (D)	reported active infected (J)	removed accumulated	removed per day	removed per day/ reported active infected	p = T=	0,5 7
				171				
march	11	582	2039	238	67	0,0329		
march	12	869	2871	275	37	0,0129		
march	13	2086	4906	326	51	0,0104		
march	14	1159	5678	713	387	0,0682		
march	15	1597	7177	811	98	0,0137		
march	16	1954	9070	872	61	0,0067		
march	17	1884	10265	1561	689	0,0671	X	Y
march	18	2943					11311,5	1019,5
march	19	3308					13711,5	1435,5
march	20	3494					15635	2453
march	21	3925					17722	2839
march	22	3272					19942,5	3588,5
march	23	6368					22987	4535
march	24	6922					27713	5132,5
		β_1	0,21817		β_2	0,10483		
		$\beta_1 T$	1,52719		$R_{0,s}$	∞		

TABLE 1. Spanish data during the March outbreak of the pandemic. $\beta_1 T > 1$ implies an infinite value of $R_{0,s}$, which means that a symptomatic causes infinitely many secondary symptomatic cases due to the spread of the virus through the asymptomatic population. The values in columns X and Y are obtained according to equation (18) in the main text.

Data source: <https://www.worldometers.info/coronavirus/country/spain/>

$$\beta_1 = \frac{\overline{Y^2} \overline{XD} - \overline{XY} \overline{YD}}{\overline{X^2} \overline{Y^2} - \overline{XY}^2},$$

$$\beta_2 = \frac{\overline{X^2} \overline{YD} - \overline{XY} \overline{XD}}{\overline{X^2} \overline{Y^2} - \overline{XY}^2},$$

where the overline notation means the average of the values in the vector.

In Table 1 Spanish data are collected from two weeks beginning on March 11th and we assume $T = 7$ (days) and $p = 0.5$ (this choice is based on global clinical information [18, 19], not related to data used here). During this period it appears that the number of infections caused by asymptomatic were so large that $\beta_1 T$ was larger than 1 and consequently $R_{0,s}$ was infinite.

A few days later (see Table 2 where data between March 21 and April 3 are considered), with strict quarantine measures (which started on March 14th but were re-inforced on March 28th) the situation seemed completely different since $R_{0,s}$ turned out to be already less than 1, as computed using (16). Here γ_2 was estimated as an average value of the quotient between the number of reported removed individuals per day (recovery plus deaths) and the total number of active infected individuals. This gives an average duration of the disease (i.e. the symptomatic phase) of 22 days. According to our previous comments on different types of parameters, a more intrinsic value (obtained from clinical observations) could be used.

Notice that $R_{0,s}$ and also the solutions of the characteristic equation $f(\lambda) = 1$ do not depend separately on the values of the parameters p and β_2 but only on their product. This allows to compute $R_{0,s}$ (and the real root of the characteristic equation) without a priori knowledge of p , just using directly J_{j-T} instead of Y_j , thus estimating via the least squares method the values of β_1 and $\beta_2 p$.

		daily new cases (D)	reported active infected (J)	removed accumulated	removed per day	removed per day/ reported active infected	p = T=	0,5 7
				2681				
march	21	4750	21990	3506	825	0,0375		
march	22	4113	24421	4347	841	0,0344		
march	23	7687	29470	5666	1319	0,0448		
march	24	8041	35273	6785	1119	0,0317		
march	25	9686	40501	9014	2229	0,0550		
march	26	10637	46406	11380	2366	0,0510		
march	27	11048	51224	14495	3115	0,0608	X	Y
march	28	7516					57345	10995
march	29	6875					60109	12210,5
march	30	7846					61569,5	14735
march	31	7967					61612	17636,5
april	1	8195					60829,5	20250,5
april	2	6120					57825,5	23203
april	3	7472					53779	25612
					γ_2	0,0450		
					$1/\gamma_2$	22,2033	days	
	β_1		0,12219		β_2	0,0120		
	$\beta_1 T$		0,85535		$R_{0,s}$	0,923		

TABLE 2. Spanish data during the lockdown. Notice that after the lockdown (which started on 15/03/20) the value of β_1 is half the previous one, which added to a much smaller value of β_2 , imply (using equation (16) in the main text) a value of $R_{0,s}$ which is already smaller than 1. The values in columns X and Y are obtained according to equation (18) in the main text. Data source: <https://www.worldometers.info/coronavirus/country/spain/>

Nevertheless, a separate estimation of both parameters is necessary for practical purposes. Indeed, let us go back to Section 2 and let us assume that $\gamma_1 \equiv 0$ and $f(\hat{\lambda}) = 1$. Then (8) implies that an exponential solution of (3) is of the form

$$(i(\tau, t), J(t)) = y(0)e^{-\hat{\lambda}t} \left(e^{-\hat{\lambda}\tau}, p \frac{e^{-\hat{\lambda}T}}{\hat{\lambda} + \gamma_2} \right),$$

where $\hat{\lambda}$ only depends on the product $\beta_2 p$. But p appears alone multiplying the second component (the symptomatic population), the knowledge of which is of crucial importance for the sanitary system.

6. DISCUSSION

More often than not, the underlying variables that determine the dynamics of a population are not directly measurable due to technical limitations. This entails the dilemma on how to choose a good model to describe the observations. The more detailed a model is, the more close it is to reality. However, a too detailed model could be useless to make predictions if the available data is not enough to estimate properly its parameters. This phenomenon was elegantly showed in [17] by considering a complex model to be the reality, generating partial data according to this model, fitting the “real” complex model and a “wrong” simpler model with that data, and finally showing that the more accurate predictions were given by the simpler model due to the large discrepancy between the real parameters of the original complex model and the ones obtained with the fit.

The spread of a pathogen is a clear example in which the available data is limited. Indeed, unless wide screening tests are performed, data consist solely on those individuals manifesting symptoms and reporting them to the health system. Asymptomatic individuals are in some sense hidden and cannot be quantified. To address the importance of the unreported cases in the spread of a disease, in this article we have analysed a model

suited for epidemics in which the prevalence of asymptomatic individuals can be large. Conditions guaranteeing its expansion or decay are given in terms of general functional parameters. Afterwards, a particular class of functional parameters, namely constant functions, has been used to fit the model with Spanish data of Covid-19.

We have found the transmission rate of asymptomatic individuals times the duration of the asymptomatic period, namely $\beta_1 T$, to be a critical parameter of the model. This value being larger than one means that the pathogen is spreading unnoticedly because each infected individual causes, on average, more than one other infection during the asymptomatic phase. Our results suggest that this was the case in Spain before the lockdown started on 14th march, since according to the data $\beta_1 T \simeq 1.5$. The model also indicates that, once the lockdown was well established, not only $\beta_1 T$ was smaller than one but also the reproduction number, namely $R_{0,s}$, was. Specifically the data give $R_{0,s} \simeq 0.9$. This means that during the symptomatic phase an infected individual caused, on average, less than one secondary symptomatic case (was it directly or indirectly through the asymptomatic population), so that the spread of the disease was controlled thanks to the lockdown.

The reproduction number derived in this article represents an alternative indicator to the ones currently used to determine if the epidemic spans or shrinks. In order to be more confident on the results derived from this indicator, however, our knowledge on the parameters p and T should be improved. Moreover, a better understanding of the infection process would allow to work with more realistic models. For example, the model presented here could be generalized to allow asymptomatic individuals to become symptomatic at different ages of infection: instead of assuming that all asymptomatic become symptomatic or cure T units of time after infection, a functional parameter would determine the rate at which this process occurs depending on the age of infection. An equation for the reproduction number in this more general case can be obtained following the same approach, but the simplifications to estimate the parameters should be different.

REFERENCES

- [1] J. Arino, S. Portet. A simple model for COVID-19. *Infectious Disease Modelling* **2020**, 5, 309–315.
- [2] J. Arino, F. Brauer, P. van den Driessche, J. Watmough, J. Wu. A model for influenza with vaccination and antiviral treatment. *Journal of Theoretical Biology* **2008**, 1, 118–130.
- [3] C. Barril, À. Calsina, J. Ripoll. A practical approach to R_0 in continuous-time ecological models. *Mathematical Methods in the Applied Sciences* **2017**, 41 (18), 8432–8445.
- [4] C. Barril, À. Calsina, J. Ripoll. On the reproduction number of a gut microbiota model. *Bulletin of Mathematical Biology* **2017**, 79 (11), 2727–2746.
- [5] C. Barril, À. Calsina, S. Cuadrado, J. Ripoll. On the basic reproduction number in continuously structured populations. *Mathematical Methods in the Applied Sciences* **2020**, <https://doi.org/10.1002/mma.6787>, 1–14.
- [6] D. Breda, F. Florian, J. Ripoll, R. Vermiglio. Efficient numerical computation of the basic reproduction number for structured populations. *Journal of Computational and Applied Mathematics* **2021**, 384, 113165.
- [7] R. H. Chisholm, P. T. Campbell, Y. Wu, S. Y. C. Tong, J. McVernon, N. Geard. Implications of asymptomatic carriers for infectious disease transmission and control. *Royal Society Open Science* **2018**, 5, 172341.
- [8] K. L. Cooke. Models for endemic infections with asymptomatic cases. I. One group. *Mathematical Modelling* **1982**, 3 no. 1, 1–15.
- [9] J.M. Cushing, O. Diekmann. The many guises of R_0 (a didactic note). *Journal of Theoretical Biology* **2016**, 404, 295–302.
- [10] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* **1990** 28 (4), 365–382.
- [11] J.A.P. Heesterbeek. A brief history of R_0 and a recipe for its calculation. *Acta biotheoretica* **2002** 50 (3), 189–204.
- [12] K.-J. Engel, R. Nagel. *A short course on operator semigroups*, Springer, 2006.
- [13] H. Inaba, H. Nishiura. The state-reproduction -number for a multistate class age structured epidemic system and its application to the asymptomatic transmission model. *Mathematical Biosciences* **2008**, 216, 77–89.
- [14] H. Inaba. *Age-structured population dynamics in demography and epidemiology*. Springer, Singapore, 2017.
- [15] W. O. Kermack, A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London* **1927**, 115, 700–721.
- [16] Z. Liu, P. Magal, O. Seydi, G. Webb. A COVID-19 epidemic model with latency period. *Infectious Disease Modelling* **2020**, 5, 323–337.
- [17] D. Ludwig, C. J. Walters. Are age-structured models appropriate for catch-effort data?. *Canadian Journal of Fisheries and Aquatic Sciences* **1985**, 42, 1066–1072.
- [18] D. P. Oran, E. J. Topol. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. *Annals of internal medicine* **2020**, 173 (5), 362–365.

- [19] J. Qin and C. You, Q. Lin T. Hu, S. Yu, X-H. Zhou. Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. *Science advances* **2020**, *6* (33), eabc1202.
- [20] H. Thieme. Spectral bound and reproduction number for infinite-dimensional population structure and time heterogeneity. *SIAM Journal on Applied Mathematics* **2009**, *70* (1), 188–211.