IMA Journal of Applied Mathematics (2022) Page 1 of 15
doi:10.1093/imamat/xxx000

Asymptotic solutions of the SIR and SEIR models well above the epidemic threshold

Gregory Kozyreff
Physics Department, Université libre de Bruxelles (U.L.B.), CP 231,
Campus de la Plaine, 1050 Bruxelles, Belgium.

[Received on 23 March 2022]

A simple and explicit expression of the solution of the SIR epidemiological model of Kermack and McKendrick is constructed in the asymptotic limit of large basic reproduction numbers R_0. The proposed formula yields good qualitative agreement already when $\mathsf{R}_0 \geqslant 3$ and rapidly becomes quantitatively accurate as larger values of R_0 are assumed. The derivation is based on the method of matched asymptotic expansions, which exploits the fact that the exponential growing phase and the eventual recession of the outbreak occur on distinct time scales. From the newly derived solution, an analytical estimate of the time separating the first inflexion point of the epidemic curve from the peak of infections is given. Finally, we use the same method on the SEIR model and find that the inclusion of the "exposed" population can dramatically alter the time scales of the outbreak.

Keywords: SIR model, SEIR model, matched asymptotic expansions, large reproduction number.

1. Introduction

The COVID-19 pandemic has impacted all aspects of our daily lives. It has thus triggered an extraordinary world-wide response across all fields of science, from virology and pharmacology to sociology. Within the scientific crowd, applied mathematicians and theoretical physicists also brought their contribution, notably by using their mathematical expertise in epidemiological modelling Bagal et al. (2020); Roda et al. (2020); Calafiore et al. (2020); Simha et al. (2020); Casares & Khan (2020); Chen et al. (2020); Chowdhury et al. (2020); Gatto et al. (2020); Bastos & Cajueiro (2020); Giordano et al. (2020); Oliveira et al. (2021); Merbis (2021). Deterministic compartment models allow one in principle to develop an accurate global view of the contagion dynamics within large populations. Compartments can be used to divide the population into age categories as well as according to the evolution of the illness, for those who have been infected Brauer et al. (2012); Gatto et al. (2020); Giordano et al. (2020). These multiple levels of description are useful and necessary, and can be made to closely reproduce the data. However, they contain a plethora of fitting parameters and can therefore be complex to interpret. Fortunately, it turns out that the simplest of all compartment models, SIR, faithfully reproduces the global dynamics at the level of a country or a large city with COVID-19 Bastos & Cajueiro (2020); Kozyreff (2021). As a result, the SIR model remains a useful tool at such global level of description. Besides, epidemiologic thinking can be used, either with the SIR model or variants thereof, to model the spread of rumours Daley & Kendall (1964), songs Rosati et al. (2021) or singer popularity Tweedle & Smith (2012) as well as fashion Acerbi et al. (2012) and fads on the internet Bauckhage et al. (2013), making the SIR model an active subject of research, nearly a century after having been put forward

© The author 2022. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

by Kermack & McKendrick (1927). We write it as follows

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI/N,$$
$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI/N - \gamma I,$$
$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I,$$

where S, I, and R respectively denote the number of susceptible, infected and removed individuals, with constant sum $N = S + I + R$. The population R includes both those who have recovered from the illness and those who have died. Above, β is the contact rate, γ is the recovery rate, and their ratio

$$\mathsf{R}_0 = \beta/\gamma$$

is the basic reproduction number. Despite its long history, there continues to be many efforts to try and solve it analytically Brauer (2005); Khan et al. (2009); Harko et al. (2014); Bohner et al. (2019); Barlow & Weinstein (2020); Giubilei (2020); Kröger & Schlickeiser (2020); Mieghem (2020); Prodanov (2020); de Souza (2021); Kudryashov et al. (2021). In their seminal paper, Kermack and McKendrick offered an approximate solution in the limit of a "small epidemic", *i.e.* when the reproduction number R_0 is just above 1. So far, this has remained the only explicit approximate analytical solution available, in the sense that it can be justified as an asymptotic limit of the true solution. Well above the epidemic threshold, Kendall showed in 1956 that the solution can be written in parametric form Kendall (1956). Unfortunately, it is only given implicitly, such that time is expressed in terms of one of the dynamical variables through an integral [see Eq. (2.9) below.] Hence, this solution is not directly interpretable and its analysis can be quite involved Mieghem (2020). Alternatives have been proposed in the form of converging series Khan et al. (2009); Barlow & Weinstein (2020), but these can involve expansions with as many as 15 up to 60 terms and, hence, are again impractical for analysis. More down-to-earth is the approach by which the solution is fitted by a well-chosen ansatz with a few parameters Clark et al. (2020); de Souza (2021). However, these ansatzes are not derived from the model and their parameters must be fitted to the data, rather than deduced from the model. During the first wave of the COVID-19 pandemic, and before confinement measures were imposed, values of R_0 between 3 and more than 6 have been estimated Nadeau et al. (2021); Liu et al. (2020) and recovery times on the order of two weeks for patients admitted in hospital have been estimated Kozyreff (2021), suggesting $\gamma = 1/15$ day^{-1} (for the first wave of the pandemic). In other context, much larger values of R_0 have been inferred Rosati et al. (2021).

1.1 *Statement of results*

The aim of this paper is to show that

$$\frac{R(t)}{N} \equiv n_r(t) \sim \begin{cases} \frac{1}{\mathsf{R}_0-1}\ln\left[ae^{(\beta-\gamma)t} + 1 - a\right], & t < t_*, \\ \frac{1}{\mathsf{R}_0-1}\ln\left[1 + \frac{1-a}{a}e^{-(\beta-\gamma)t}\right] + n_{r,\infty} - e^{-\gamma(t-t_*)}, & t > t_*, \end{cases} \tag{1.1}$$

where a is the infected fraction of the population at $t = 0$, while $n_{r,\infty}$ and t_* are given by

$$n_{r,\infty} \sim 1 - e^{-\mathsf{R}_0\left(1-e^{-\mathsf{R}_0}\right)}, \qquad t_* \sim \frac{n_{r,\infty} - 1}{\gamma} - \frac{\ln a}{\beta - \gamma}. \tag{1.2}$$

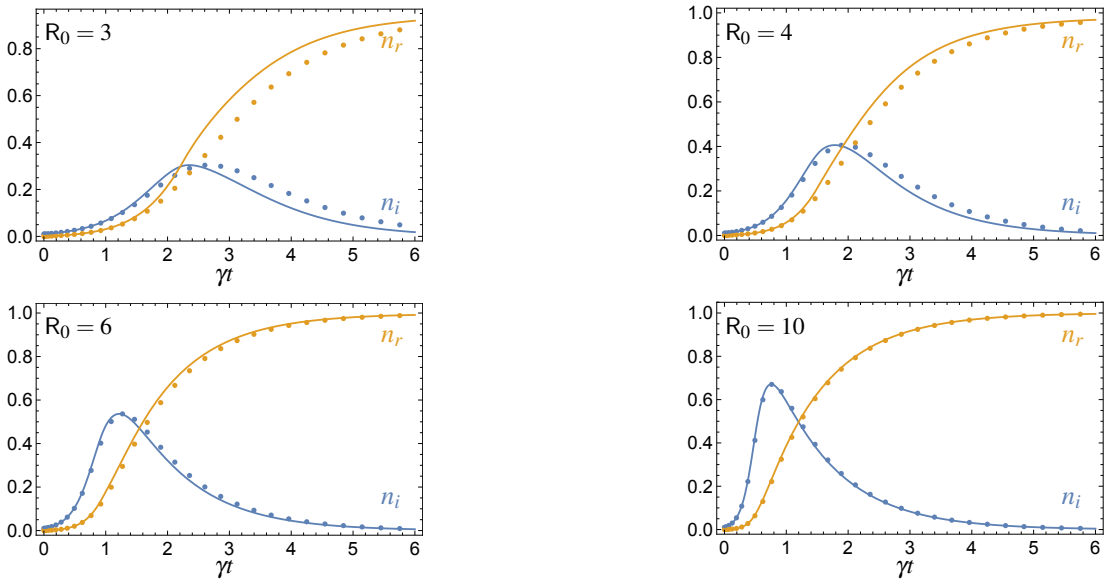

FIG. 1. Comparison between the approximate formulas (1.1) and (1.3) (full lines) and the numerical solution of the SIR equation (dotted lines) for several values of the reproduction number R_0. $n_i(0) = a = 0.01$. Note that, once R_0 is given, the parameter γ only appears through the dimensionless product γt.

Given $n_r(t)$, one may infer the infected fraction of the population exactly as

$$\frac{I(t)}{N} \equiv n_i(t) = 1 - n_r(t) - (1-a)\, e^{-\mathsf{R}_0 n_r(t)}. \tag{1.3}$$

Eq. (1.1) is remarkably simple and compact. The "$\sim$" sign in it indicates that it holds asymptotically in the limit $\mathsf{R}_0 \gg 1$. However, a good qualitative agreement with the numerical solution is already found for $\mathsf{R}_0 = 3$ and the approximate formula rapidly becomes as good as exact with increasing R_0, see Fig. 1. Hence, Eq. (1.1) complements the classical "small epidemic" formula derived by Kermack & McKendrick (1927), which is valid in the limit $0 < \mathsf{R}_0 - 1 \ll 1$. When, in 1956, D. G. Kendall presented the implicit analytical solution [Eq. (2.9) below] of the SIR model, he noted:

> "*It is curious that the Kermack and McKendrick approximation should have been accepted without comment for nearly thirty years; the exact solution is easily obtained and the difference between the two can be of practical significance.*" (Kendall, 1956, p. 150)

Similarly, the reader will find that the derivation of the present asymptotic formula is quite simple and could have been established much earlier with well-established asymptotic techniques.

As an illustration of the usefulness of the present analytical expression, let us compute the time elapsed between two important moments of the epidemic, namely the time t_1 of first inflexion of the curve $n_i(t)$ and the time t_2 of the peak of the epidemic. In appendix, we show that the former happens when n_r is equal to n_1, solution of

$$\mathsf{R}_0 \frac{\mathsf{R}_0 e^{-\mathsf{R}_0 n_1}}{1 - \mathsf{R}_0 e^{-\mathsf{R}_0 n_1}} = \frac{1 - \mathsf{R}_0 e^{-\mathsf{R}_0 n_1}}{1 - n_1 - e^{-\mathsf{R}_0 n_1}}, \tag{1.4}$$

while the latter happens when

$$n_r = n_2 = \frac{\ln \mathsf{R}_0}{\mathsf{R}_0}, \tag{1.5}$$

in the limit $a \to 0$. It turns our that $t_1 < t_*$ and that $t_2 > t_*$. Hence, using Eq. (1.1), we have

$$\gamma t_1 \sim \frac{\ln(1/a)}{\mathsf{R}_0 - 1} + \frac{1}{\mathsf{R}_0 - 1} \ln\left[e^{(\mathsf{R}_0-1)n_1} - 1\right]. \tag{1.6}$$

Meanwhile, for t_2, we have

$$n_2 = \frac{1}{\mathsf{R}_0 - 1} \ln\left[1 + \frac{1-a}{a} e^{-(\beta-\gamma)t_2}\right] + n_{r,\infty} - e^{-\gamma(t_2 - t_*)}. \tag{1.7}$$

Above, we may assume in first approximation that t_2 is sufficiently large that the logarithmic term is negligible. Thus, we obtain the more manageable expression

$$n_2 \sim n_{r,\infty} - e^{-\gamma(t_2-t_*)}, \qquad \to \gamma t_2 \sim \gamma t_* + \ln\left(n_{r,\infty} - n_2\right), \tag{1.8}$$

and find

$$\gamma(t_2 - t_1) \sim n_{r,\infty} - 1 + \ln\left(n_{r,\infty} - n_2\right) - \frac{1}{\mathsf{R}_0 - 1} \ln\left[e^{(\mathsf{R}_0-1)n_1} - 1\right]. \tag{1.9}$$

A comparison between this approximate formula and the numerics is given in Fig. 2, suggesting that the relative error associated to Eq. (1.9) is $O\left[(\mathsf{R}_0 \ln \mathsf{R}_0)^{-1}\right]$. Collecting data in real-time allows one to detect when the slope of the epidemic curve has reached a maximum; $t_2 - t_1$ is then the time left before the epidemic starts to recede. In a context such as the first wave of the COVID-19 pandemic, being able to predict when the epidemic will reach its peak can be a valuable asset both to manage health resources and to communicate to the public.

2. Preliminaries

Before embarking into the asymptotic analysis of the SIR model, we first recall a few well-known facts. Eliminating S thanks to the relation $S(t) + I(t) + R(t) = N$, the evolution equations for the fractions $n_i = I/N$ and $n_r = R/N$ are

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = (\beta - \gamma)\, n_i - \beta\,(n_i + n_r)\, n_i, \tag{2.1}$$

$$\frac{\mathrm{d}n_r}{\mathrm{d}t} = \gamma n_i. \tag{2.2}$$

They should be solved with the initial condition

$$n_i(0) = a \ll 1, \qquad n_r(0) = 0. \tag{2.3}$$

At the start of the outbreak, $n_i, n_r \ll 1$ and the linearised evolution equations immediately yield

$$n_i \approx a e^{(\beta-\gamma)t}, \qquad n_r \approx \frac{a}{\mathsf{R}_0 - 1} e^{(\beta-\gamma)t}. \tag{2.4}$$

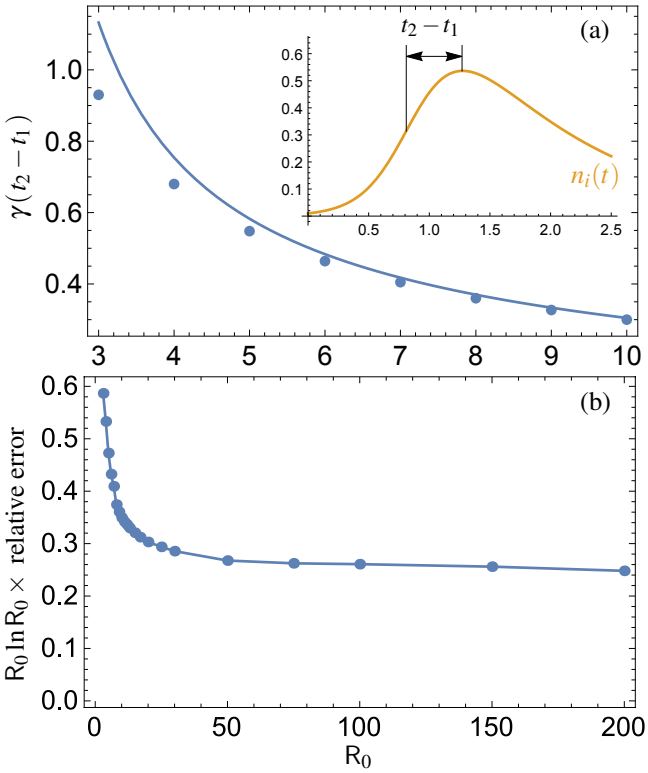

FIG. 2. Time from inflexion to epidemic maximum as a function of R_0. (a) Comparison between numerical solution (dots) and asymptotic formula (full line). (b) Plot of the relative error $|y_{\text{num}}/y_{\text{asy}} - 1|$ multiplied by $R_0 \ln R_0$, where y_{num} and y_{asym} designate the numerical value and asymptotic approximation of $t_2 - t_1$, respectively.

(Here, the '$\approx$' sign is used because nonlinear terms in Eqs. (2.1) and (2.2) are neglected.) On the other hand, dividing Eq. (2.1) by Eq. (2.2), one obtains

$$\frac{\mathrm{d}}{\mathrm{d}n_r}\left(n_i+n_r-1\right)=-\mathsf{R}_0\left(n_i+n_r-1\right), \tag{2.5}$$

which directly yields Eq. (1.3). Thanks to the explicit dependence $n_i[n_r(t)]$, Eq. (2.2) eventually becomes a first-order nonlinear ODE:

$$\frac{\mathrm{d}n_r}{\mathrm{d}t}=\gamma\left[1-n_r(t)-(1-a)\,e^{-\mathsf{R}_0 n_r(t)}\right]. \tag{2.6}$$

As $t\to\infty$, n_r tends to the value $n_{r,\infty}$ that makes the right hand side vanish:

$$n_{r,\infty}=1-(1-a)\,e^{-\mathsf{R}_0 n_{r,\infty}}\approx 1-e^{-\mathsf{R}_0 n_{r,\infty}}. \tag{2.7}$$

In the large-R_0 limit, the second term in the right hand side above is small, so that a recursive resolution can be set up. Note that, since $a\ll 1$, as is generically the case, we neglect it in (2.7). An excellent approximation is given by

$$n_{r,\infty}\sim 1-e^{-\mathsf{R}_0\left(1-e^{-\mathsf{R}_0}\right)}, \tag{2.8}$$

with an absolute error of less than 0.002 for $\mathsf{R}_0>3$. Alternatively, one may express $n_{r,\infty}$ in terms of the tabulated Lambert function Mieghem (2020).

It is immediate to see that Eq. (2.6) can formally be integrated as

$$\gamma t=\int_0^{n_r}\frac{\mathrm{d}n}{1-n-(1-a)\,e^{-\mathsf{R}_0 n}}. \tag{2.9}$$

This expression has been known and studied for a long time Kendall (1956). Note that if both γ and β are functions of time, such that their ratio $\mathsf{R}_0=\beta/\gamma$ remains constant, then Eq. (2.9) remains valid, with γt replaced by the effective time $\tau=\int\gamma(t)\mathrm{d}t$ Kröger & Schlickeiser (2020). Recently, a thorough analysis of the integrand of Eq. (2.9) was carried out, yielding upper and lower bounds for the function $t(n_r)$ Mieghem (2020). In addition, series representations of the integral based either on Taylor expansion of the integrand near the origin or based on its poles in the complex-n plane were given in that work. The first of these approaches effectively amounts to assume that n is small and thus generalizes Kermack and McKendrick approximate solution. However, the formulas in Mieghem (2020) are rather involved, making the inversion of the relation $t(n_r)$ laborious.

3. Matched asymptotic expansions

The method consists in deriving separate approximations at the early stage and at the later stage of the outbreak. One the one hand, the outer solution has a characteristic time γ^{-1} and describes the decaying phase of the epidemic. On the other hand, the inner solution applies to the growing phase and evolves on a shorter time scale. In the language of asymptotics, the $\mathsf{R}_0\gg 1$ limit produces a *boundary layer* problem Bender & Orszag (1999). The two approximations are constrained by matching conditions at intermediate stages of the outbreak. Thanks to the fact that they have an overlapping range of validity, one can then construct a composite solution, Eq. (1.1), that is uniformly valid in time.

3.1 *Outer solution*

When $n_r = O(1)$, $e^{-\mathsf{R}_0 n_r}$ is exponentially small and can at first be neglected in Eq. (2.6). We thus have

$$\frac{\mathrm{d}n_r}{\mathrm{d}t} \sim \gamma(1-n_r)\,, \qquad \to n_r \sim n_r^{(o)} = 1 - e^{-\gamma(t-t_o)}, \tag{3.1}$$

where the superscript (o) refers to "outer" approximation and t_o is an arbitrary constant. However, knowing the actual limiting value of n_r as $t \to \infty$, a better approximation is

$$n_r^{(o)} = n_{r,\infty} - e^{-\gamma(t-t_o)}, \tag{3.2}$$

where $n_{r,\infty}$ is given in Eq. (2.8). Since $n_{r,\infty}$ and 1 only differ by an exponentially small quantity, they are asymptotically equivalent as $\mathsf{R}_0 \to \infty$. However, the correction brought about by using $n_{r,\infty}$ in Eq. (3.2) considerably improves the approximation for moderately large R_0 and thus expands the range of usability of the final result, Eq. (1.1).

Note that $n_r^{(o)}$ is only defined for $t > t_o$, since n_r must be a positive number.

3.2 *Inner solution*

At earlier stages of the epidemics, n_r is $O\left({\mathsf{R}_0}^{-1}\right)$. More precisely, Eq. (2.4) suggests to write

$$n \sim n_r^{(i)}(t) = \frac{\nu(t)}{\mathsf{R}_0 - 1}, \tag{3.3}$$

where (i) refers to "inner" approximation. Then Eq. (2.6) becomes

$$\frac{\mathrm{d}\nu}{\mathrm{d}t} = (\mathsf{R}_0 - 1)\,\gamma\left[1 - e^{-\frac{\mathsf{R}_0}{\mathsf{R}_0-1}\nu}(1-a) - \frac{1}{\mathsf{R}_0-1}\nu\right], \tag{3.4}$$

$$\sim (\beta-\gamma)\left[1 - e^{-\nu}(1-a)\right]. \tag{3.5}$$

Unlike Eq. (2.6), this last equation can be integrated explicitly:

$$\frac{e^{\nu}}{e^{\nu}-1+a}\frac{\mathrm{d}\nu}{\mathrm{d}t} \sim \beta-\gamma, \qquad \to \ln\left(\frac{e^{\nu}-1+a}{a}\right) \sim (\beta-\gamma)\,t, \tag{3.6}$$

where we used the fact that $\nu(0) = 0$. Hence,

$$\nu \sim \ln\left[a e^{(\beta-\gamma)t} - 1 + a\right], \tag{3.7}$$

and

$$n_r^{(i)} \sim \frac{1}{\mathsf{R}_0 - 1}\ln\left[a e^{(\beta-\gamma)t} - 1 + a\right]. \tag{3.8}$$

Just as we used $n_{r,\infty}$ instead of 1 in the outer solution, we here and henceforth use $1/(\mathsf{R}_0 - 1)$ instead of $1/\mathsf{R}_0$: the two are asymptotically equivalent to leading order in R_0, but the former yields superior numerical approximation, on account of the linear expression (2.4).

3.3 *Matching*

So far, we have derived two approximations that efficiently describe the solution at different moments of the epidemic. However, the constant t_o in Eq. (3.2) is still unknown. To determine it, we must impose that (3.2) and (3.8) match in an intermediate region. Loosely speaking, one wants that

$$\lim_{t\to\infty} n_r^{(i)}(t) \sim \lim_{t\to 0} n_r^{(o)}(t). \tag{3.9}$$

To formalise this statement more precisely, let us define the matching region as

$$t = t_* + \eta\Delta t, \tag{3.10}$$

where η is such that $\gamma\eta\Delta t \ll 1$, while $(\beta-\gamma)\,\eta\Delta t \gg 1$. This double asymptotic constraint is possible since $(\beta-\gamma)/\gamma = \mathsf{R}_0 - 1 \gg 1$ by assumption. On the one hand, we have

$$n_r^{(o)}(t_* + \eta\Delta t) = n_{r,\infty} - e^{-\gamma(t_*-t_o)-\gamma\eta\Delta t}, \tag{3.11}$$

$$\sim n_{r,\infty} - e^{\gamma(t_*-t_o)} + e^{-\gamma(t_*-t_o)}\gamma\eta\Delta t. \tag{3.12}$$

On the other hand, the inner solution yields

$$n_r^{(i)} \sim \frac{1}{\mathsf{R}_0-1}\ln\left[ae^{(\beta-\gamma)t} - 1 + a\right], \tag{3.13}$$

$$= \frac{1}{\mathsf{R}_0-1}\left(\ln\left[ae^{(\beta-\gamma)t}\right] + \ln\left[1 - \frac{1-a}{a}e^{-(\beta-\gamma)t}\right]\right), \tag{3.14}$$

$$= \frac{1}{\mathsf{R}_0-1}\left\{(\beta-\gamma)t + \ln a + \ln\left[1 - \frac{1-a}{a}e^{-(\beta-\gamma)t}\right]\right\}, \tag{3.15}$$

$$= \gamma t + \frac{\ln a}{\mathsf{R}_0-1} + \frac{1}{\mathsf{R}_0-1}\ln\left[1 - \frac{1-a}{a}e^{-(\beta-\gamma)t}\right], \tag{3.16}$$

$$\sim \gamma t_* + \gamma\eta\Delta t + \frac{\ln a}{\mathsf{R}_0-1}. \tag{3.17}$$

Equating expressions (3.12) and (3.17), we obtain the two conditions

$$n_{r,\infty} - e^{-\gamma(t_*-t_o)} = \gamma t_* + \frac{\ln a}{\mathsf{R}_0-1}, \qquad \text{and} \qquad e^{-\gamma(t_*-t_o)} = 1. \tag{3.18}$$

Hence, we have $t_o = t_*$ and

$$\gamma t_* = n_{r,\infty} - 1 - \frac{\ln a}{\mathsf{R}_0-1}. \tag{3.19}$$

3.4 *Composite solution*

As indicated above, $n_r^{(o)}$ is only defined for $t > t_o = t_*$. For $t < t_*$, only the inner solution holds:

$$n_r(t < t_*) \sim n_r^{(i)}. \tag{3.20}$$

On the other hand, for $t > t_*$, we may construct a uniformly valid approximation as

$$n_r(t > t_*) \sim n_r^{(i)} + n_r^{(o)} - (\text{common part}), \tag{3.21}$$

where the common part is the time-dependance that $n_r^{(i)}$ and $n_r^{(o)}$ have in common in the matching region, and which should be removed to avoid double counting. One has, from (3.17),

$$\text{(common part)} = \gamma t_* + \gamma\eta\Delta t + \frac{\ln a}{\mathsf{R}_0 - 1} = \gamma t + \frac{\ln a}{\mathsf{R}_0 - 1}. \tag{3.22}$$

Hence

$$\begin{aligned} n_r(t > t_*) &\sim \frac{1}{\mathsf{R}_0 - 1} \ln\left[a e^{(\beta-\gamma)t} - 1 + a\right] + n_{r,\infty} - e^{-\gamma(t-t_*)} - \gamma t - \frac{\ln a}{\mathsf{R}_0 - 1} \\ &= \frac{1}{\mathsf{R}_0 - 1} \ln\left[1 - \frac{1-a}{a} e^{-(\beta-\gamma)t}\right] + n_{r,\infty} - e^{-\gamma(t-t_*)}. \end{aligned} \tag{3.23}$$

Eqs. (3.20) and (3.23) yield Eq. (1.1), which completes the derivation.

4. Extension to the SEIR model

The method of matched asymptotic expansions can be applied to more elaborate models as well. We demonstrate this on the SEIR model, which extends the SIR by adding a population of contaminated individuals that are not contagious yet. This population is called "exposed" and designated by E. Once a person is exposed to the disease, it takes an average time γ_e^{-1} before he or she falls into the category I of infected people. The SEIR model reads as follows

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI/N, \qquad \frac{\mathrm{d}E}{\mathrm{d}t} = \beta SI/N - \gamma_e E, \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \gamma_e E - \gamma I, \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I. \tag{4.1}$$

As before, $S(t)+E(t)+I(t)+R(t) = N$, allowing us to eliminate $S(t)$. Defining the fractions $n_e = E/N$, $n_i = I/N$ and $n_r = R/N$, we thus obtain

$$\frac{\mathrm{d}n_e}{\mathrm{d}t} = \beta\left(1 - n_e - n_i - n_r\right) n_i - \gamma_e n_e, \qquad \frac{\mathrm{d}n_i}{\mathrm{d}t} = \gamma_e n_e - \gamma n_i, \qquad \frac{\mathrm{d}n_r}{\mathrm{d}t} = \gamma n_i. \tag{4.2}$$

We propose to solve it with the initial conditions $n_i(0) = a \ll 1$, $n_e(0) = n_r(0) = 0$. The introduction of the new variable n_e comes with a new parameter γ_e. We discard the situation $\gamma_e \gg \beta, \gamma$, which corresponds to the limit of a vanishing duration of the "exposed" state. In that limit, it is easy to find that $\gamma_e n_e \sim \beta\left(1 - n_i - n_r\right)$, which leads us back to the SIR model. Rather, we focus on the limit $\beta \gg \gamma_e, \gamma$ and assume that γ_e is larger than, but of the same order of magnitude as, γ. Following exactly the same reasoning as in Sec. 2, one easily derives the first integral

$$n_e(t) + n_i(t) = 1 - n_r(t) - (1-a)\, e^{-\mathsf{R}_0 n_r(t)}, \tag{4.3}$$

where $\mathsf{R}_0 \gg 1$ is still defined as the ratio β/γ.

4.1 *Outer region*

In the outer region, $n_r = O(1)$ and Eq. (4.3) yields $n_r \sim n_r^{(o)} = 1 - n_e^{(o)} - n_i^{(o)}$. We are thus left with

$$\frac{\mathrm{d}n_e^{(o)}}{\mathrm{d}t} \sim -\gamma_e n_e^{(o)}, \qquad \frac{\mathrm{d}n_i^{(o)}}{\mathrm{d}t} \sim \gamma_e n_e^{(o)} - \gamma n_i^{(o)}. \tag{4.4}$$

The solution is

$$n_e^{(o)} \sim B e^{-\gamma_e(t-t_o)}, \qquad n_i^{(o)}(t) \sim A e^{-\gamma(t-t_o)} - B e^{-\gamma_e(t-t_o)}. \tag{4.5}$$

where t_o, A, and B are constants to be determined by matching.

4.2 *Inner region*

In the inner region, we find the appropriate units and time scale by writing

$$n_e \sim n_e^{(i)}(t) = v_e(\tau), \qquad n_i \sim n_i^{(i)}(t) = u_i v_i(\tau), \qquad n_r \sim n_r^{(i)}(t) = u_r v_r(\tau), \qquad \tau = \sigma t, \tag{4.6}$$

where the constants u_i, u_r and σ are arbitrary. Eqs. (4.2) become

$$\frac{\mathrm{d}v_e}{\mathrm{d}\tau} = \frac{\beta u_i}{\sigma}\left(1 - v_e - u_i v_i - u_r v_r\right) v_i - \frac{\gamma_e}{\sigma} v_e, \qquad \frac{\mathrm{d}v_i}{\mathrm{d}\tau} = \frac{\gamma_e}{\sigma u_i} v_e - \frac{\gamma}{\sigma} v_i, \qquad \frac{\mathrm{d}v_r}{\mathrm{d}\tau} = \frac{\gamma u_i}{\sigma u_r} v_i. \tag{4.7}$$

Let us choose σ, u_i and u_r such that $\beta u_i/\sigma = 1$, $\gamma_e/(\sigma u_i) = 1$ and $\gamma u_i/(\sigma u_r) = 1$, that is

$$\sigma = \sqrt{\beta \gamma_e}, \qquad u_i = \sqrt{\gamma_e/\beta} \ll 1, \qquad u_r = \gamma/\beta = \mathsf{R}_0{}^{-1} \ll 1. \tag{4.8}$$

Hence, to leading order, the first two of Eqs. (4.7) reduce to

$$\frac{\mathrm{d}v_e}{\mathrm{d}\tau} \sim \left(1 - v_e\right) v_i, \qquad \frac{\mathrm{d}v_i}{\mathrm{d}\tau} \sim v_e, \tag{4.9}$$

where we neglected $O\left(\sqrt{\gamma_e/\beta}\right)$ and $O\left(\mathsf{R}_0{}^{-1}\right)$ terms. Equivalently, we have

$$\frac{\mathrm{d}^2 v_i}{\mathrm{d}\tau^2} = \left(1 - \frac{\mathrm{d}v_i}{\mathrm{d}\tau}\right) v_i. \tag{4.10}$$

The above equation should be solved with the initial conditions

$$v_i(0) = a\sqrt{\beta/\gamma_e} \equiv \alpha, \qquad \frac{\mathrm{d}v_i}{\mathrm{d}\tau}(0) = v_e(0) = 0. \tag{4.11}$$

There is only one parameter left in this problem: α. Let us denote the numerical solution of this problem as

$$v_i = \phi(\tau, \alpha). \tag{4.12}$$

It does not seem possible to express this function in closed form. However, since it depends only on τ and α, the initial condition, its numerical determination is very simple and represents a progress compared to the complete problem. Furthermore, it is easy to see that the asymptotic behaviour of $v_i(\tau)$ as $\tau \to \infty$ is

$$v_i \sim \tau - \tau_*(\alpha), \tag{4.13}$$

as illustrated in Fig. 3. The function $\tau_*(\alpha)$ is a universal function that can be determined numerically once and for all. A good numerical fit, with absolute error of less than 0.013 in the range $0 < \alpha < 2.5$, is given by

$$\tau_*(\alpha) \approx \ln\alpha - 0.287 + 0.417\alpha + 0.073\alpha^2. \tag{4.14}$$

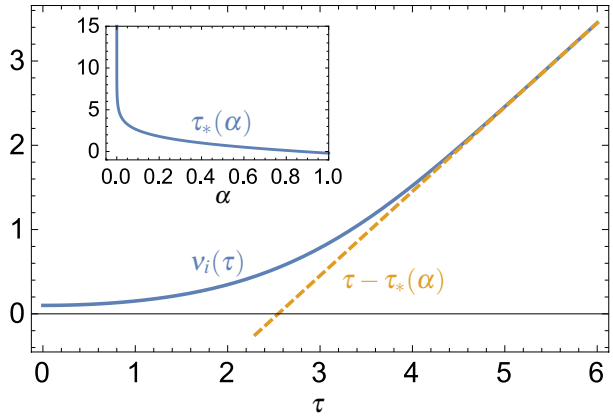

FIG. 3. Solution of Eq. (4.10) with $v_i(0) = \alpha$, $v_i'(0) = 0$, and $\alpha = 0.1$. Dashed line, asymptotic behaviour of $v_i(\tau)$. Inset: the function $\tau_*(\alpha)$.

4.3 *Matching*

As $\tau \to \infty$, we have, combining Eqs. (4.8) and (4.13):

$$n_i^{(i)} \sim \gamma_e \left[t - \frac{\tau_* \left(a\sqrt{\beta/\gamma_e} \right)}{\sqrt{\gamma_e \beta}} \right]. \tag{4.15}$$

On the other hand, as $t \to t_o$, the outer solution yields

$$n_i^{(0)} \sim A - B + (B\gamma_e - A\gamma)(t - t_o). \tag{4.16}$$

The matching conditions are, therefore,

$$A = B = \frac{\gamma_e}{\gamma_e - \gamma}, \qquad t_o = \frac{\tau_* \left(a\sqrt{\beta/\gamma_e} \right)}{\sqrt{\gamma_e \beta}}. \tag{4.17}$$

The common part between the inner and outer solutions is simply $\gamma_e(t - t_o)$. We now have all the elements to construct a uniformly valid composite asymptotic approximation of $n_i(t)$.

For $t < t_o$, only the inner solution exists because the outer solution yields negative values of n_i:

$$n_i(t < t_o) \sim n_i^{(i)}(t). \tag{4.18}$$

For $t > t_o$, the composite solution is

$$n_i(t > t_o) \sim n_i^{(i)}(t) + n_i^{(0)}(t) - (\text{common part}). \tag{4.19}$$

Eventually, we obtain

$$n_i(t) \sim \sqrt{\frac{\gamma_e}{\beta}}\, \phi\left(\sqrt{\beta\gamma_e}\, t, a\sqrt{\beta/\gamma_e} \right) + H(t - t_o) \left\{ \frac{\gamma_e}{\gamma_e - \gamma} \left[e^{-\gamma(t-t_o)} - e^{-\gamma_e(t-t_o)} \right] - \gamma_e(t - t_o) \right\}, \tag{4.20}$$

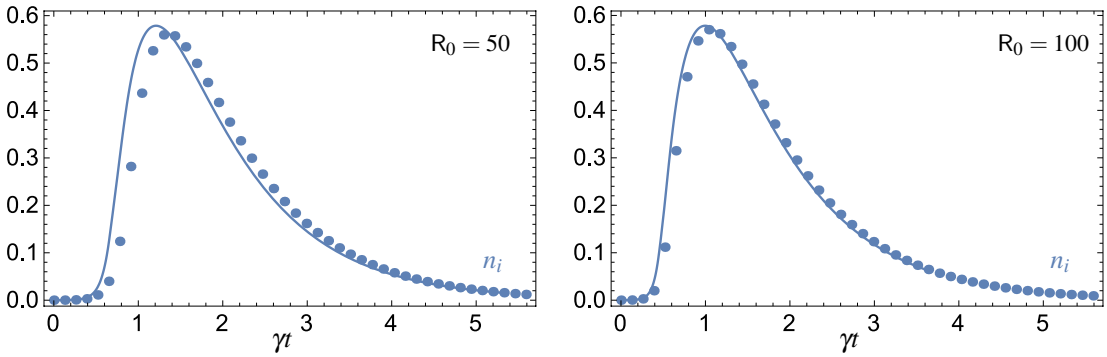

FIG. 4. Comparison between the approximate formulas (4.20) (full lines) and the numerical solution of the SEIR equation (dotted lines) for large values of the reproduction number R_0. $n_i(0) = a = 0.0001$, $n_r(0) = 0$, $n_e(0) = 0$. The time unit is chosen so that $\gamma = 1$, $\gamma_e = 3$ and $\beta = \mathsf{R}_0$.

where $H(t)$ is the Heaviside step function.

A comparison between these asymptotic approximations and the numerical solution is given in Fig. 4. Note that a much larger value of R_0 is required to obtain a good quantitative agreement. This is because β and, hence, R_0 always appear in square roots. Hence, contrarily to the SIR model, the small parameter of the problem is not $1/\mathsf{R}_0$ but $1/\sqrt{\mathsf{R}_0}$. Moreover, the fast scale of the problem is not βt anymore but $\sqrt{\beta\gamma_e}\,t$. It is remarkable that the mere introduction of an intermediate variable such as the exposed population $E(t)$ completely modifies the times scales of the dynamics, even though γ_e is of the same order of magnitude as γ.

5. Conclusion

We have shown that a compact and explicit approximate solution of the SIR model can be derived in the limit of large R_0. The interest of this solution is that it is already effective when the reproduction number is only moderately large. Given that SIR is in itself a crude approximation of reality, it makes little sense from an applied perspective to look for an approximation that is exact to within 1%. Taking the time between inflexion and maximum of the epidemic curve as a measure of accuracy, the present theory, Eq. (1.9) yields an answer with less than 10% error for R_0 just above 4. With influenza, a reproduction number exceeding the value of 3 appears rare but possible White et al. (2009), as was the case with the 1918 influenza pandemic Mills et al. (2004). In the case of the COVID-19 pandemic estimates yield R_0 in the range 1.5 to 6.49 Liu et al. (2020) and particularly around 3 in France Nadeau et al. (2021) and between 3.5 and 4 in Korea and in the Hubei province (China) during the first wave Choi & Ki (2020). For measles, R_0 can be on the order of 10 or even larger Guerra et al. (2017). Hence, the large-R_0 limit considered here is relevant, even though, of course, the assumption of a constant ratio β/γ does not apply to the COVID-19 pandemic. While the formulas derived in this paper appear accurate enough, it is to be noted that they can be improved. Only leading-order expressions have been used for the inner and outer solutions that apply to the ascending and receding phases of the outbreak, respectively. The inner solution can be improved by the keeping $O\left(1/(\mathsf{R}_0 - 1)\right)$ corrections in Eq. (3.4) and treating them as perturbations. The same can be done with $O\left(e^{-\mathsf{R}_0 n_r}\right)$ terms in the equation for the outer solution. Even within the present theory, a correction to t_2 could be computed by perturbation, thus improving Eq. (1.9). In this paper, we applied the technique of matched asymptotic expansions to the SIR model.

This technique could in principle be applied to more general epidemiological models, leading to simple and reliable expressions in the limit of large reproduction numbers. We demonstrated this point with the SEIR model. Interestingly, the analysis shows that the mere introduction of the population E of "exposed", not yet contagious, individuals profoundly alters the times scales of the outbreak. Also, in this second example the effective small parameter is ${\mathsf{R}_0}^{-1/2}$ instead of ${\mathsf{R}_0}^{-1}$ and so the error committed in retaining only the leading order approximation is $O\left({\mathsf{R}_0}^{-1/2}\right)$.

Aknowledgement

G.K. is a Research Associate with the Belgian Fonds de la Recherche Scientifique (FNRS).

A. Derivation of formulas (1.4) and (1.5)

From Eq. (2.1), $\mathrm{d}n_i/\mathrm{d}t$ vanishes when $n_i = 1 - \frac{1}{\mathsf{R}_0} - n_r$. By substitution into (1.3), one immediately obtains (1.5).

Next, let us enquire when the second time derivative of n_i vanishes. Writing $n_i = n_i\left[n_r(t)\right]$, let us denote

$$n_i' = \frac{\mathrm{d}n_i}{\mathrm{d}n_r}. \tag{A.1}$$

Then,

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = n_i'\frac{\mathrm{d}n_r}{\mathrm{d}t}, \tag{A.2}$$

$$\frac{\mathrm{d}^2 n_i}{\mathrm{d}t^2} = n_i''\left(\frac{\mathrm{d}n_r}{\mathrm{d}t}\right)^2 + n_i'\frac{\mathrm{d}^2 n_r}{\mathrm{d}t^2} = n_i'\left(\frac{\mathrm{d}n_r}{\mathrm{d}t}\right)^2\left[\frac{n_i''}{n_i'} + \left(\frac{\mathrm{d}n_r}{\mathrm{d}t}\right)^{-2}\frac{\mathrm{d}^2 n_r}{\mathrm{d}t^2}\right], \tag{A.3}$$

and one thus seek for the condition

$$\frac{n_i''}{n_i'} = -\left(\frac{\mathrm{d}n_r}{\mathrm{d}t}\right)^{-2}\frac{\mathrm{d}^2 n_r}{\mathrm{d}t^2}. \tag{A.4}$$

Dividing (2.1) by (2.2), one has

$$n_i' = \mathsf{R}_0 - 1 - \mathsf{R}_0\left(n_i + n_r\right). \tag{A.5}$$

Hence, $n_i'' = -\mathsf{R}_0\left(n_i' + 1\right)$ and

$$\frac{n_i''}{n_i'} = -\mathsf{R}_0\left(\frac{n_i'+1}{n_i'}\right) = -\mathsf{R}_0\left(\frac{\mathsf{R}_0 - \mathsf{R}_0\left(n_i+n_r\right)}{\mathsf{R}_0 - 1 - \mathsf{R}_0\left(n_i+n_r\right)}\right)$$
$$= \mathsf{R}_0\frac{\mathsf{R}_0\left(1-a\right)e^{-\mathsf{R}_0 n_r}}{1-\mathsf{R}_0\left(1-a\right)e^{-\mathsf{R}_0 n_r}} \sim \mathsf{R}_0\frac{\mathsf{R}_0 e^{-\mathsf{R}_0 n_r}}{1-\mathsf{R}_0 e^{-\mathsf{R}_0 n_r}}, \tag{A.6}$$

as $a \to 0$. On the other hand, neglecting a compared to 1, (2.6) yields

$$\frac{\mathrm{d}^2 n_r}{\mathrm{d}t^2} = -\gamma\frac{\mathrm{d}n_r}{\mathrm{d}t}\left(1 - \mathsf{R}_0 e^{-\mathsf{R}_0 n_r}\right). \tag{A.7}$$

Hence,

$$-\left(\frac{\mathrm{d}n_r}{\mathrm{d}t}\right)^{-2}\frac{\mathrm{d}^2 n_r}{\mathrm{d}t^2} = \gamma\left(\frac{\mathrm{d}n_r}{\mathrm{d}t}\right)^{-1}\left(1 - \mathsf{R}_0 e^{-\mathsf{R}_0 n_r}\right) = \frac{1-\mathsf{R}_0 e^{-\mathsf{R}_0 n_r}}{1 - n_r - e^{-\mathsf{R}_0 n_r}}. \tag{A.8}$$

Equating (A.6) with (A.8) yields (1.4).

REFERENCES

Acerbi, A., Ghirlanda, S. & Enquist, M. (2012) The Logic of Fashion Cycles. *PLoS ONE*, **7**, e32541.

Bagal, D. K., Rath, A., Barua, A. & Patnaik, D. (2020) Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods. *Chaos, Solitons & Fractals*, **140**, 110154.

Barlow, N. S. & Weinstein, S. J. (2020) Accurate closed-form solution of the SIR epidemic model. *Physica D: Nonlinear Phenomena*, **408**, 132540.

Bastos, S. B. & Cajueiro, D. O. (2020) Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil. *Scientific Reports*, **10**, 19457.

Bauckhage, C., Kersting, K. & Hadiji, F. (2013) Mathematical models of fads explain the temporal dynamics of internet memes. In *Seventh international AAAI conference on weblogs and social media*.

Bender, C. M. & Orszag, S. A. (1999) *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory (v. 1)*. Springer.

Bohner, M., Streipert, S. & Torres, D. F. (2019) Exact solution to a dynamic SIR model. *Nonlinear Analysis: Hybrid Systems*, **32**, 228–238.

Brauer, F. (2005) The Kermack–McKendrick epidemic model revisited. *Mathematical Biosciences*, **198**, 119–131.

Brauer, F., Castillo-Chavez, C. & Castillo-Chavez, C. (2012) *Mathematical models in population biology and epidemiology*, volume 2. Springer.

Calafiore, G. C., Novara, C. & Possieri, C. (2020) A time-varying SIRD model for the COVID-19 contagion in Italy. *Annual Reviews in Control*, **50**, 361–372.

Casares, M. & Khan, H. (2020) A dynamic model of COVID-19: Contagion and implications of isolation enforcement. Technical report, Carleton Economics Working Papers (CEWP). Carleton University.

Chen, Y., Cheng, J., Jiang, Y. & Liu, K. (2020) A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification. *Journal of Inverse and Ill-posed Problems*, **28**, 243–250.

Choi, S. & Ki, M. (2020) Estimating the reproductive number and the outbreak size of COVID-19 in Korea. *Epidemiology and Health*, **42**, e2020011.

Chowdhury, R., , Heng, K., Shawon, M. S. R., Goh, G., Okonofua, D., Ochoa-Rosales, C., Gonzalez-Jaramillo, V., Bhuiya, A., Reidpath, D., Prathapan, S., Shahzad, S., Althaus, C. L., Gonzalez-Jaramillo, N. & Franco, O. H. (2020) Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. *European Journal of Epidemiology*, **35**, 389–399.

Clark, D. E., Welch, G. & Peck, J. S. (2020) Approximate relationships between SIR and logistic models. *medRxiv*.

Daley, D. J. & Kendall, D. G. (1964) Epidemics and rumours. *Nature*, **204**, 1118–1118.

de Souza, F. M. C. (2021) An Approximate Analytical Solution to the SIR Model. *Available at SSRN: https://ssrn.com/abstract=3780708 or http://dx.doi.org/10.2139/ssrn.3780708*.

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R. & Rinaldo, A. (2020) Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, **117**, 10484–10491.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Filippo, A. D., Matteo, A. D. & Colaneri, M. (2020) Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, **26**, 855–860.

Giubilei, R. (2020) Closed form solution of the SIR model for the COVID-19 outbreak in Italy. *medRxiv*.

Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y. & Crowcroft, N. S. (2017) The basic reproduction number (R_0) of measles: a systematic review. *The Lancet Infectious Diseases*, **17**, e420–e428.

Harko, T., Lobo, F. S. & Mak, M. (2014) Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*, **236**, 184–194.

Kendall, D. G. (1956) Deterministic and stochastic epidemics in closed populations. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob*, volume 4, pages 149–165.

Kermack, W. O. & McKendrick, A. G. (1927) A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A*, **115**, 700–721.

Khan, H., Mohapatra, R. N., Vajravelu, K. & Liao, S. (2009) The explicit series solution of SIR and SIS epidemic models. *Applied Mathematics and Computation*, **215**, 653–669.

Kozyreff, G. (2021) Hospitalization dynamics during the first COVID-19 pandemic wave: SIR modelling compared to Belgium, France, Italy, Switzerland and New York City data. *Infectious Disease Modelling*, **6**, 398–404.

Kröger, M. & Schlickeiser, R. (2020) Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor. *Journal of Physics A: Mathematical and Theoretical*, **53**, 505601.

Kudryashov, N. A., Chmykhov, M. A. & Vigdorowitsch, M. (2021) Analytical features of the SIR model and their applications to COVID-19. *Applied Mathematical Modelling*, **90**, 466–473.

Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklv, J. (2020) The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, **27**(2).

Merbis, W. (2021) Exact epidemic models from a tensor product formulation. *arXiv preprint arXiv:2102.11708*.

Mieghem, P. V. (2020) Exact solution of the Kermack and McKendrick SIR differential equations. *arXiv:2010.03253*.

Mills, C. E., Robins, J. M. & Lipsitch, M. (2004) Transmissibility of 1918 pandemic influenza. *Nature*, **432**, 904–906.

Nadeau, S. A., Vaughan, T. G., Scire, J., Huisman, J. S. & Stadler, T. (2021) The origin and early spread of SARS-CoV-2 in Europe. *Proceedings of the National Academy of Sciences*, **118**(9), e2012008118.

Oliveira, J. F., Jorge, D. C. P., Veiga, R. V., Rodrigues, M. S., Torquato, M. F., da Silva, N. B., Fiaccone, R. L., Cardim, L. L., Pereira, F. A. C., de Castro, C. P., Paiva, A. S. S., Amad, A. A. S., Lima, E. A. B. F., Souza, D. S., Pinho, S. T. R., Ramos, P. I. P. & Andrade, R. F. S. (2021) Mathematical modeling of COVID-19 in 14.8 million individuals in Bahia, Brazil. *Nature Communications*, **12**, 333.

Prodanov, D. (2020) Analytical Parameter Estimation of the SIR Epidemic Model. Applications to the COVID-19 Pandemic. *Entropy*, **23**(1), 59.

Roda, W. C., Varughese, M. B., Han, D. & Li, M. Y. (2020) Why is it difficult to accurately predict the COVID-19 epidemic?. *Infectious Disease Modelling*, **5**, 271–281.

Rosati, D. P., Woolhouse, M. H., Bolker, B. M. & Earn, D. J. D. (2021) Modelling song popularity as a contagious process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **477**(2253), 20210457.

Simha, A., Prasad, R. V. & Narayana, S. (2020) A simple stochastic sir model for COVID-19 infection dynamics for Karnataka: Learning from Europe. *arXiv preprint arXiv:2003.11920*.

Tweedle, V. & Smith, R. J. (2012) A mathematical model of Bieber Fever: The most infectious disease of our time. *Understanding the dynamics of emerging and re-emerging infectious diseases using mathematical models. Transworld Research Network*.

White, L. F., Wallinga, J., Finelli, L., Reed, C., Riley, S., Lipsitch, M. & Pagano, M. (2009) Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza and Other Respiratory Viruses*, **3**(6), 267–276.