A new SEIR epidemic model with applications to the theory of eradication and control of diseases, and to the calculation of $R_0$

Sen-Zhong Huang

Institut für Mathematik, Universität Rostock, Universitätsplatz 1, D-18055 Rostock, Federal Republic of Germany

ARTICLE INFO

Article history:
Received 3 October 2007
Received in revised form 5 June 2008
Accepted 11 June 2008
Available online 24 June 2008

Keywords:
SEIR epidemic model
Basic reproductive number
Thresholds

ABSTRACT

We present a novel SEIR (susceptible–exposure–infective–recovered) model that is suitable for modeling the eradication of diseases by mass vaccination or control of diseases by case isolation combined with contact tracing, incorporating the vaccine efficacy or the control efficacy into the model. Moreover, relying on this novel SEIR model and some probabilistic arguments, we have found four formulas that are suitable for estimating the basic reproductive numbers $R_0$ in terms of the ratio of the mean infectious period to the mean latent period of a disease. The ranges of $R_0$ for most known diseases, that are calculated by our formulas, coincide very well with the values of $R_0$ estimated by the usual method of fitting the models to observed data.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction and overview of results

$R_0$, the so-called basic reproductive (reproduction) number of a communicable disease, defined as the expected number of secondary infectious cases generated by one typical primary case in an entirely susceptible and sufficiently large population (Anderson and May [1]), is a key for controlling the epidemic spread. For instance, the general guideline for mass vaccination is that the herd immunity of a population against the disease must be above some critical immunization threshold $q_c$ that is a function of $R_0$ and some other factors like vaccine efficacy. A very simple but illustrative choice of such a critical immunization threshold is $q_c = 1 - 1/R_0$, used in the traditional theory of epidemiology [1] under the unrealistic assumption (Benenson [3], CDC [6]) that the vaccine is perfectly effective. Although epidemiologists and theoreticians ([3],6], Halloran et al. [24,25], Longini et al. [41]) have been aware of this for a long time, no confident relation between the critical immunization threshold $q_c$ and the vaccine efficacy was drawn. It gives rise to the need of establishing more realistic epidemic models for the mass vaccination theory.

A further motivation for establishing new epidemic models is the estimation of the basic reproductive number $R_0$. As shown by the very recent case of the 2002–2003 epidemic SARS (Donnelly et al. [15], Lipsitch et al. [36], Riley et al. [53], Bauch et al. [2]), an accurate estimation of $R_0$ is essential for a successful control of newly emerging diseases by the public health control measures. A large number of compartmental (both deterministic and stochastic) epidemic models (Anderson and May [1], Hethcote [28], Diekmann and Heesterbeek [14]) have been proposed for this purpose (see Heffernan et al. [27] for a comprehensive review). But most of these models rely on the traditionally used assumption that both transmission processes of latency and recovery states are exponentially distributed. This assumption is biologically unrealistic, because it corresponds to assuming that such a transmission process is memoryless in the sense that the chance of change from a state ‘A’ to another state ‘B’ in a given time interval is independent of the time since the entrance into the previous state ‘A’. More realistic distributions are, for example, the gamma or the log-normal distributions (Sartwell [54]). However, the inclusion of gamma distributions, as shown by Lloyd [38,39] and Wearing et al. [59], could change drastically the patterns and dynamics of the epidemic models, as well as the estimate of the basic reproductive number $R_0$.

In this work, we will present (in Section 2) a new SEIR (susceptible–exposure–infective–recovered) epidemic model, which covers the classical SEIR model (Hethcote [28]) as a special case and has the following advantage than most of the known mathematical epidemic models: our new SEIR model is suitable for the aim of incorporating epidemiological factors like the vaccine efficacy and the varying of transmission patterns into the model settings. Hence, it is reasonable to expect that our new SEIR model fits better to reality.

Below we highlight the following three results obtained by our new SEIR model: a novel theory for the eradication of diseases by mass vaccination, a refined theory for the control of newly emerging diseases by the public health control measures of isolating infectious individuals combined with contact tracing and quarantining suspected individuals, and four formulas for calculating...
the basic reproductive number $R_0$. For simplicity, we assume as usual that the population under discussion is homogeneously mixed.

### 1.1. Eradication of diseases by vaccination

Let $V_{Es}$ be the vaccine efficacy against susceptibility, and $V_{Ei}$ the vaccine efficacy against infectiousness (Halloran et al. [24,25], Longini et al. [41]). Moreover, we assume that the vaccination is so effective that the following inequality is satisfied:

$$R_0(1 - V_{Es})(1 - V_{Ei}) < 1. \quad (1.1a)$$

Let $f_e$ and $0 < f_t < 1$, be the solution of the following quadratic equation

$$R_0(1 - f_eV_{Es})(1 - f_tV_{Ei}) = 1. \quad (1.1b)$$

Then $f_t$ is a critical immunization threshold; that is, if the coverage $f$ of the herd immunity in a homogeneously mixed population exceeds $f_t$, then it can be expected that the disease will be eradicated by the mass vaccination, no matter how complicated the transmission patterns of the disease could be.

### 1.2. Control of outbreaks

In the absence of effective vaccine or treatment, the symptom-based public control measures follow the following baselines (Fracer et al. [21]): (i) effective isolation of symptomatic individuals and (ii) tracing and quarantining the contacts of symptomatic cases. Let $\theta$ be the proportion of asymptomatic infectious whose transmission occurs prior to symptoms. Let $a_e$ be the efficacy of isolation/quarantine and $a_t$ be the efficacy of contact tracing. (We assume that the duration of isolation/quarantine is long enough so that the infection is cleared and no longer infectious.) The value $a_t$ could equally be thought of as the probability that an infected individual will be isolated immediately after he/she becomes symptomatic, and the value $a_e$ could equally be thought of as the probability that an asymptomatic infective will be first detected by contact tracing and then isolated. Let $\sigma = \sigma t$ (with $\delta \in (0, 1]$) be the mean time of delay between isolation and the starting point of infection for those asymptomatic infectives who are detected by contact tracing. Here $\sigma$ is the mean infectious period for those infectives who are neither isolated nor detected by contact tracing. If the control measures are such that the following inequality

$$(1 - \theta)a_t + \theta(1 - \delta)a_t > 1 - 1/R_0 \quad (1.2)$$

is satisfied, then our theory ensures that the outbreak will be brought under control, no matter how complicated the transmission patterns of the disease could be. (We remark that formulas similar to (1.2) have been obtained previously by Müller et al. [45] and Fracer et al. [21] using different methods.)

This result indicates in particular that a newly emerging disease could be contained by isolation combined with contact tracing and quarantine alone, provided that these control measures are sufficiently effective when compared to the basic reproductive number $R_0$ of the disease. The newest example for this is the control of the 2002–2003 epidemic SARS (see e.g., Donnelly et al. [15], Lipsitch et al. [36]).

The above theory also confirms the usually employed strategy for controlling an epidemic: containing the infection sources (infectives), controlling the epidemic.

### 1.3. Estimation of $R_0$

As indicated above, a core for a successful eradication or control of diseases is the estimation of the basic reproductive number $R_0$. In our present work, we follow the usual guideline [27,28] used for both the classical SIR model of Kermack and McKendrick [33] and the classical SEIR model [28] by defining $R_0$ as the product of the intrinsic mean infection rate (denoted by $\beta$) and the mean infectious period (denoted by $\gamma$), i.e., $R_0 := \beta \times \gamma$. The usual way to estimate $R_0$ is to estimate both $\beta$ and $\gamma$ by fitting observed data from known outbreaks into corresponding models; see, e.g., [1,7,9,11,13,16,18,36,43].

As the disease spreads, the mean infectious period $\gamma$ of a disease can be directly estimated, by using a few data, with or without models. However, estimation of the intrinsic mean infection rate $\beta$ depends strongly on the model and the fruitfulness of data used and is in general very subtle; cf. [1,7,9,11,13,16,18,36,43]. Our present method for estimating $\beta$ is purely theoretical and relies on our new SEIR model. Based on some sophisticated probabilistic arguments (cf. Section 4), our method yields that the intrinsic mean infection rate $\beta$ can be determined in terms of the three main factors of a disease: the mean latent period (denoted by $\tau$), the mean infectious period, and the transmission patterns of the latency and recovery processes. The typical transmission patterns of latency and recovery will be divided into four types: light (Type I), moderate (Type II), severe (Type III), extremely severe (Type IV). Our method yields the following four formulas (1.4)–(1.7) for calculating the basic reproductive numbers $R_{01}^{IV}(z)$ associated with the four types of transmission patterns, in terms of the ratio $z := \text{mean infectious period} : \text{mean latent period}$. (1.3)

- $R_0^{I}(z)$, the minimal reproductive number associated with transmission patterns of Type I (with the slowest latency process and the fastest recovery process).
  $$R_0^{I}(z) = (1 + y(z)z) \exp(y(z)) \quad \text{with}$$
  $$y(z) := \frac{z}{(1 + 2z + \sqrt{1 + 4z})} \quad (1.4)$$

- $R_0^{II}(z)$, the middle reproductive number associated with transmission patterns of Type II (with the mean latency and recovery processes).
  $$R_0^{II}(z) = \frac{1}{8} \left(2 + (1 + 4z)y^2(z)\right)^2 \quad (1.5)$$

- $R_0^{III}(z)$, the maximal reproductive number associated to transmission patterns of Type III (with the fastest latency process and the slow recovery process).
  $$R_0^{III}(z) = (1 + \sqrt{20z}(1 + \sqrt{20z}/(1 + 20z))) \quad (1.6)$$

- $R_0^{IV}(z)$, the largest reproductive number associated with transmission patterns of Type IV (with the fastest latency process and the extremely slow recovery process).
  $$R_0^{IV}(z) = (1 + \sqrt{100z}(1 + \sqrt{100z}/(1 + 100z))) \quad (1.7)$$

All four functions $R_0^{1-IV}(z)$ are strictly increasing functions of $z = \gamma/\tau$, the ratio of the mean infectious period $\gamma$ to the mean latent period $\tau$, and they satisfy the following inequalities:

$$R_0^{I}(z) < R_0^{II}(z) < R_0^{III}(z) < R_0^{IV}(z) \quad (z > 0). \quad (1.8)$$

These four formulas will be used as follows: according to the type of its transmission patterns, we assign one of the main categories ‘mild’ and ‘severe’ to the disease and then estimate its basic reproductive number.

#### 1.3.1. Category ‘mild’

We assume that diseases of the category ‘mild’ have transmission patterns which are between the light type (Type I) and the
moderate type (Type II), and the superspreading events (SSEs) generated by a disease of the category ‘mild’ correspond to the case that the transmission patterns are of the severe type (Type III). Assuming that a disease of the category ‘mild’ has a mean latent period ranging from $k$ to $l$ units of time (e.g., days), and a mean infectious period ranging from $m$ to $n$ units of time, then the basic reproductive numbers $R_0$ and the reproductive number $R_0^\text{SSE}$ generated by SSEs for this ‘mild’ disease are estimated by

$$R_0^\text{m}(m/l) < R_0 < R_0^\text{SSE}(n/k), \quad R_0^\text{SSE} = R_0^\text{m}(n/k). \quad (1.9)$$

### 1.3.2. Category ‘severe’

We assume that diseases of the category ‘severe’ have transmission patterns which are between the severe type (Type III) and the extremely severe type (Type IV), and the superspreading events (SSEs) generated by a disease of the category ‘severe’ correspond to the case that the transmission patterns are of the extremely severe type (Type IV). Assuming that a disease of the category ‘severe’ has a mean latent period ranging from $k$ to $l$ units of time (e.g., days), and a mean infectious period ranging from $m$ to $n$ units of time, then the basic reproductive numbers $R_0$ and the reproductive number $R_0^\text{SSE}$ generated by SSEs for this ‘severe’ disease are estimated by

$$R_0^\text{s}(m/l) < R_0 < R_0^\text{SSE}(n/k), \quad R_0^\text{SSE} = R_0^\text{s}(n/k). \quad (1.10)$$

In applying the above method to most known diseases, our classification is as follows: The category ‘mild’ includes the mild diseases like Hepatitis B, Polio, Scarlet fever and HIV (Aids) which have a long infectious period (when compared to their mean latent periods) but have an observed small basic reproductive number $R_0$. The newly (re)-emerging diseases SARS, Ebola, AHC, FMD, influenza, influenza pandemic and avian influenza are put also into the category ‘mild’, since they have an observed relatively small basic reproductive number $R_0$. The category ‘severe’ includes the rest severe and extremely severe diseases like Chickenpox, Mumps, Rubella, Measles, etc. which have an observed large basic reproductive number $R_0$. The newly emerging infections Acute HIV-1 and Acute SIV belong also to this category.

Although our method for estimating $R_0$ is purely theoretical, the resulting estimates of $R_0$ (given in Table 2 of Section 3 and Table 4 of Section 4) fits very well to the published estimates of $R_0$ (see, e.g., [1,7,9,11,13,16,18,36,43]) using data from known outbreaks and using other models. This indicates that the four functions $R_0^\text{m}(z)$ given above can be used as empirical formulas for estimating $R_0$ for unknown diseases – a point that is especially valuable for the control of a newly emerging disease, because the three main factors (the mean latent period, the mean infectious period and the types of the transmission patterns), which are needed for estimating the values of $R_0$ for that disease, are observable as the disease spreads.

The present study can be extended in several directions. These include explicit incorporation of the heterogeneity of the population (cf. Anderson and May [1], Hethcote [28]), the spatial and temporal varying of the transmission patterns, the control measures as well as the population growth. Some examples are presented in Appendix D. We point out that this framework can also be applied to the eradication and control of network viruses.

Further content of this article is organized as follows: in Section 2 we present the new SEIR model. In Section 3 we establish the aforementioned novel theory for the eradication and control of diseases. In Section 4 we describe in details our method for estimating $R_0$ and present the numerical results. In the appendix section we give the detailed mathematical proofs of the results mentioned in the previous sections.

### 2. A new SEIR model and a new threshold theorem

We consider the situation that an epidemic with a mean latent period $\tau$ and a mean infectious period $\sigma$ emerges and spreads in a homogeneously mixed population for which time-dependent control measures (including vaccination, isolation and contact tracing, etc.) have been implemented during the epidemic.

The new SEIR model and its derivation: The dynamics of such an epidemic can be described by the following novel susceptible–exposure–infective–recovered (SEIR) model $S \rightarrow E \rightarrow I \rightarrow R$:

$$S'(t) = -(r(t)/N)(g(t) - R(t))S(t),$$
$$F(t) = N - S(t),$$
$$g(t) = \int_0^t F(t-s)W(s) \, ds,$$
$$R(t) = \int_0^t g'(t-s)\Lambda(t-s,s) \, ds.$$  

In the above, $N$ is the fixed total population size; $S(t)$ is the number of susceptibles at time $t$; $F(t)$ is the cumulative number of deaths that emerge within the time interval $[0,t]$; $g(t)$ is the cumulative number of infectives that emerge within $[0,t]$; and $R(t)$ is the cumulative number of infectives that are recovered (removed) within $[0,t]$.

It is clear that every infective must pass the stage of being an exposure. This implies that $g(0) = 0$ and $R(0) = 0$. The value $g_0 := F(0) = N - S(0)$ is called the initial (seed) epidemic size.

In the sequel, we assume that each of the four functions $(S(t), F(t), g(t), R(t))$ is an absolutely continuous function with an almost everywhere existing and bounded derivative. We note that the derivatives $F(t)$ and $g(t)$ can be interpreted as follows: The value of $F(t)$ (resp., $g(t)$) can be thought of as the number of new exposures (resp., new infectives) at time $t$, i.e. the number of new individuals that are infected (resp., have become being infectious) at time $t$. Similarly, the value of $R(t)$ can be thought of as the number of newly recovered infectives at time $t$. In the practical use, the value of $R(t)$ can be thought of as the number of newly detected and reported cases at time $t$.

Before further explaining things involved in the model Eq. (2.1), we point out that the new SEIR model Eq. (2.1.1) in terms of the four functions $(S(t), F(t), g(t), R(t))$ can be reformulated (though with relatively complicated representations) in terms of the four functions $(S(t), E(t), I(t), R(t))$ that come from the usual formalism of the classical SEIR model [1,14,28]. To see this, we recall that in the usual formalism of the classical SEIR model one considers four possible subsequent states for individuals in the population with a constant size $N$: susceptibility $(S)$, latency $(E)$ (i.e., infected but not yet infectious), infectiousness $(I)$ (i.e., infectious but not yet recovered), and recovery $(R)$ (recovered or removed by isolation, death, etc.). Thus, if $S(t), E(t), I(t)$ and $R(t)$ are the number of individuals at time $t$ in the corresponding $(S)(E)(I)(R)$ state, then we have

$$S(t) + E(t) + I(t) + R(t) = N. \quad (2.2a)$$

We note that in the formalism of model Eq. (2.1), both functions $S(t)$ and $R(t)$ keep their classical meaning. Moreover, the number $E(t)$ (resp., $I(t)$) of exposures (resp., infectives) at time $t$ can be calculated by

$$E(t) = F(t) - g(t), \quad I(t) = g(t) - R(t),$$

or equivalently,

$$F(t) = E(t) + I(t) + R(t), \quad g(t) = I(t) + R(t). \quad (2.2c)$$

More exactly, the first equation $E(t) = F(t) - g(t)$ in (2.2b) simply expresses that at time $t$ the number $E(t)$ of individuals that are in
the latency state is equal to the cumulative number \( F(t) \) of exposures that emerge within the time interval \([0, t]\), minus the cumulative number \( -g(t) \) of exposures that have become being infectious within \([0, t]\). Similarly, the second equation \( I(t) = g(t) - R(t) \) in (2.2b) expresses that at time \( t \) the number \( I(t) \) of active infectives is equal to the cumulative number \( g(t) \) of infectives that emerge within the time interval \([0, t]\), minus the cumulative number \( R(t) \) of infectives that are recovered within \([0, t]\).

From (2.2b) and (2.2c) we can see that our new model Eq. (2.1) is essentially a SEIR model.

We now return to the new model Eq. (2.1).

The three functions \( r, W \) and \( A \) involved in Eq. (2.1) are interpreted as follows.

The function \( r(t) \) is assumed to be bounded and piecewise continuous. At a time point \( t \geq 0 \), the value \( r(t) \) is the mean infection rate per unit time per active infective.

The function \( W \) is an absolutely continuous (a.c., for short) cumulative distribution function (CDF, for short) on \( \mathbb{R}_+ \) with expectation (mean) value \( \tau \) and zero initial derivative, i.e.,

\[
\int_0^\infty (1 - W(t)) \, dt = \tau, \quad W'(0) = 0. 
\] (2.3a)

The absolute continuity of \( W \) means that \( W \) has an almost everywhere existing derivative \( W' \) (called the probability density function of \( W \)) which is Lebesgue-integrable on \( \mathbb{R}_+ \). The requirement \( W'(0) = 0 \) corresponds to the fact that no exposed individual will immediately become infectious after infection.

For the function \( A \) we assume that for each fixed \( t \geq 0 \) the function \( A(t, s) \) in \( s \geq 0 \) is an a.c. CDF on \( \mathbb{R}_+ \) with a finite expectation

\[
\Sigma(t) := \int_0^\infty (1 - A(t, s)) \, ds < \infty. 
\] (2.3b)

The value \( \Sigma(t) \) is called effective mean infectious period. In general, \( \Sigma(t) \) depends on the time-varying control measures and is not greater than the intrinsic mean infectious period \( \sigma : \Sigma(t) \leq \sigma \quad (t \geq 0). \) (2.3c)

The functions \( W(t) \) and \( A(t, s) \) have the following interpretation. The value \( 1 - W(t) \) is the probability that an exposed individual will remain in the exposure state after its emergence of \( t \) units of time. For each fixed \( t \geq 0 \), the value \( A(t, s) \) is the probability that an infective individual that emerges at \( t \) will be recovered at the later time point \( t + s \), i.e., after an emergence period of \( s \) units of time. Hence, we call \( A \) also the (time-dependent) recovery cumulative distribution function.

In this article, by a CDF (cumulative distribution function) \( P \) on \( \mathbb{R}_+ \) we mean a function that is monotone increasing, right-continuous for all \( t > 0 \) and that satisfies \( P(0) = 0 \) and \( P(t) \rightarrow 1 \) as \( t \rightarrow \infty \). Moreover, the integral with respect to CDFs is interpreted as the Lebesgue–Stieltjes integral, as commonly done in the probability theory.

To finish the model setting, we explain below how our integro-differential model Eq. (2.1) is derived.

First we do this for both Eqs. (2.1a) and (2.1b). Using (2.2b) we see that Eq. (2.1b) is purely a restatement of the classical balance Eq. (2.2a). We point out that there is a direct way to derive Eq. (2.1b). In fact, at \( t \geq 0 \), the difference \( N - S(t) \) is equal to the number of individuals that have been first exposed to the disease and then infected within the time interval \([0, t]\). This yields the balance Eq. (2.1b) stating \( N - S(t) = F(t) \), since \( F(t) \) is the cumulative number of exposures that emerge newly within the time interval \([0, t]\).

Our derivation of Eq. (2.1a) just follows the usual formalism of the classical SEIR model [1,28] by assuming that the decrease of susceptibles follows the law of mass action in the sense that

\[
S'(t) = -\lambda(t)S(t),
\] (2.4a)

where \( \lambda(t) \) is the so-called ‘force of infection’ [Anderson and May 1, p. 62]. That is, at time \( t \) the product \( \lambda(t) \Delta t \) is the probability that a given susceptible host will become infected in the small time interval \( \Delta t \). Since only those infectives that are active (i.e., infectious but not yet removed) will be able to transmit their infection, the force of infection \( \lambda(t) \) can be calculated as follows: \( N \times \lambda(t) \) is equal to the product of \( r(t) \), the time-dependent mean infection rate per unit time per active infective, with the number \( I(t) \) of infectives that are still active at time \( t \), i.e.,

\[
\lambda(t) = r(t)L(t)/N. 
\] (2.4b)

On the other hand, we have that \( I(t) = g(t) - R(t) \) by (2.2b). Hence, Eq. (2.1a) is obtained by substituting (2.4b) into (2.4a).

There is another way to understand the above derivation of Eq. (2.1a). We use the balance equation (2.1b) to find that \( F'(t) = -S'(t) \). Thus, Eq. (2.1a) can be rewritten as

\[
F'(t) = \lambda(t)S(t)
\] (2.4c)

with a ‘force of infection’ \( \lambda(t) \) that takes the form (2.4b). The new equation (2.4c) can be derived in a very direct way. In fact, we note that the number of individuals that are infected within the time interval \([t, t + \Delta t] \) is equal to the difference \( F(t + \Delta t) - F(t) \). Thus, the quotient \( (F(t + \Delta t) - F(t))/S(t) \) can be approximately thought of as the probability that a given susceptible host will become infected in the small time interval \([t, t + \Delta t] \). In terms of the ‘force of infection’ \( \lambda(t) \), this can be expressed as

\[
(F(t + \Delta t) - F(t))/S(t) \approx \lambda(t)\Delta t,
\]
or equivalently,

\[
(F(t + \Delta t) - F(t))/\Delta t \approx \lambda(t)S(t). 
\] (2.4d)

By letting \( \Delta t \rightarrow 0 \) in (2.4d), we obtain Eq. (2.4c).

Next we give the derivation of Eqs. (2.1c) and (2.1d) that describe the latency and recovery processes. The basic assumption that leads to (2.1c) and (2.1d) is:

Both latency and recovery processes follow linear laws.

We explain them in a more detailed way.

(i) For Eq. (2.1c) we assume that the latent process \( E \rightarrow I \) is linear and determined by the transmission CDF \( W \) as follows:

\[
g'(t) = F(0)W'(t) + \int_0^t F(t - s)W(s) \, ds \quad (t \geq 0). 
\] (2.5)

The term \( F(0)W'(t) \) in (2.5) represents the transmission of the initial exposures (amounting to \( F(0) \)) to new infectives at time \( t \). The integral term in (2.5) represents the transmission of the exposures that have emerged within the time interval \([0, t] \) to new infectives at \( t \). Eq. (2.1c) is the result of integrating (2.5) and using the initial condition \( g(0) = 0 \). The first equality in condition (2.3a) corresponds to the requirement that the mean latent period be \( \tau \).

(ii) Eq. (2.1d) is derived under the assumption that the transmission process \( I \rightarrow R \) for the infectives is linear and determined by the time-dependent CDFs \( A(t, \cdot) \). A more illustrative explanation is as follows: at a fixed time point \( t > 0 \), the density of infectives that emerge \( s \) units of time before \( t \) and that will be removed at \( t \) is the product \( g'(t - s) \times A(t - s, s) \). Hence, the cumulative number \( R(t) \) of removed infectives at time \( t \) is given by

\[
R(t) = \int_0^t g'(t - s)A(t - s, s) \, ds, 
\]
which is just Eq. (2.1d).

We give some comments to the new SEIR model Eq. (2.1). As seen above, both Eqs. (2.1a and 2.1b) are derived under the usual formalism for the classical SEIR model [28,1]. What are really new in our model are Eqs. (2.1c) and (2.1d) for describing the latency
and recovery processes in terms of the general CDFs W and A. We point out that many possible choices for the CDFs W and A (e.g., multi-staged exponentially distributed CDFs) that are met in practical applications, will lead our original model Eq. (2.1) to systems of ordinary differential equations (odes). Some such examples are given in Appendix D. In particular, we find that such systems of odes contain the classical SEIR model as a special case (see Appendix E).

With general choices of the CDFs W and A, the new SEIR model Eq. (2.1) is an integro-differential system. However, we can show, with some more manipulations (details given in Appendix A), that the original system Eq. (2.1) can be simplified to a single equation for the function F of the cumulative number of exposures combined with the relations for calculating the cumulative numbers g and R of infectives and removed infectives. Because the equation for F is a non-linear integral equation of the Volterra type, the usual fixed-point method (see Appendix A) is used to establish the well-posedness of the integral equation for F, and thus the well-posedness of Eq. (2.1). Particularly, it yields that for each initial epidemic size 〈0 < 〈N < N〉, the system Eq. (2.1) admits a unique, global and increasing solution F such that F(0) = 〈N and F(t) < N for all t ≥ 0.

R_eff(t), the effective reproductive number, and R_0, the basic reproductive number. The number R_eff(t) given by

\[ R_{\text{eff}}(t) := \int_0^\infty r(t + s)(1 - A(t, s)) \, ds \quad (t \geq 0) \]  

will be called the effective reproductive number at time t. The reason why we use this terminology is based on the following observation. We consider the case where prior immunity and control measures are absent. For this case we have that

r(t) = 1, \quad A(t, s) = A_0(s) \quad (s, t \geq 0),

where $\beta$ is the intrinsic mean infection rate and $A_0$ is the intrinsic recovery CDF. Correspondingly, the mean value

\[ \sigma := \int_0^\infty (1 - A_0(t)) \, dt \]  

is just the mean infectious period of the disease. Under the choices (2.7a) for r and A, the function $R_{\text{eff}}(t)$ given by (2.6) is identical to a constant $R_0$ given by

\[ R_0 := \sigma \beta. \]  

$R_0$ is called the basic reproductive number.

Our definition of $R_0$ is consistent with the general baseline (cf. Anderson and May [1], Hethcote [28]) where one defines the basic reproductive number as the expected number of secondary infectious cases generated by one typical primary case in an entirely susceptible and sufficiently large population.

The effective reproductive numbers $R_{\text{eff}}(t)$ and control of epidemic: the threshold theorem. The asymptotic behavior of solutions to Eq. (2.1) is as follows: a direct consequence of the well-posedness of Eq. (2.1) is that the epidemic course described by Eq. (2.1) will develop in such a way that

\[ R(t) \leq g(t) \leq F(t) < N \]  

for all t ≥ 0, and will be stopped eventually with a finite final size of $g_{\infty}$ infectives in such a way that

\[ F(t) + |F(t) - g_{\infty}| + |g(t) - g_{\infty}| + |R(t) - g_{\infty}| \to 0 \quad \text{as} \quad t \to \infty, \]  

and

\[ (N - F(t)) \exp \left( -\frac{1}{N} \int_0^t R_{\text{eff}}(s) g(s) \, ds + \frac{1}{N} H(t) \right) \]  

\[ = N - g_{\infty} \quad (\forall t \geq 0), \]  

where

\[ H(t) := \int_0^t g(s) \left( \int_0^{t-s} r(s + x)(1 - A(s, x)) \, dx \right) \, ds. \]  

Note that the value of $F(t)$ is the number of new exposures at time t. Hence, the statement $F(t) \to 0$ as $t \to \infty$ in (2.9b) says that the epidemic course will be stopped eventually. (Clearly, one reason for the stopping of the epidemic course is the boundedness of the total population size.) The other statement in (2.9b) stating that the three functions $F(t)$, $g(t)$ and $R(t)$ have the same limit as $t \to \infty$ restates exactly the fact that all exposures will become infectious and will eventually be removed from the disease either by death or discharge. In the literature, Eq. (2.9c) is called the final size equation and is also satisfied by many other models [28,27]. Historically, the final size equation was derived for the first time by Kermack and McKendrick [33] in 1927 for the classical SIR model.

In a realistic situation, the control measures will be implemented at a certain time point $t_0 \geq 0$ after the disease has spread. The baseline for the implementation of control measures is to reduce the effective reproductive number $R_{\text{eff}}(t)$ given by (2.6). It is expected that the epidemic will be brought under control if the effective reproductive number will be maintained below some level $\delta < 1$ after $t_0$, a certain time point after the intervention. Our following result confirms the theoretical correctness of this expectation.

**Theorem 2.1.** Let

\[ R_{\text{eff}}(t) := \int_0^\infty r(t + s)(1 - A(t, s)) \, ds \quad (t \geq 0), \]  

be the effective reproductive number and let $g_{\infty}$ be the final epidemic size satisfying Eq. (2.9b).

We have the following assertions (i)–(iii).

(i) If there exists some $t_0 \geq 0$ such that

\[ R_{\text{eff}}(t) < 1, \]  

then there holds

\[ g_{\infty} / F(t_0) < 1 / (1 - R_0). \]  

(ii) If there exists a constant $R_{\infty} > 1$ such that

\[ R_{\text{eff}}(t) \geq R_{\infty} \quad \forall t \geq 0, \]  

then there holds

\[ g_{\infty} / N > p, \]  

where $p \in (0, 1)$ is the unique solution of the following equation

\[ 1 - p = e^{-R_{\infty} p}. \]  

(iii) If there holds

\[ R_{\text{eff}}(t) \geq 1 \quad \forall t \geq 0, \]  

then

\[ g_{\infty} \geq \sqrt{g_0 N}. \]  

The proof of Theorem 2.1 will be given in Appendix B.

We remark that the so-called ‘threshold property’ of the basic reproductive number $R_0$ given by (2.7c) is interpreted as follows: In the absence of prior immunity and control measures we have that $R_{\text{eff}}(t) \equiv R_0$. Hence, if $R_0 < 1$, then Theorem 2.1 yields that the final epidemic size $g_{\infty}$ is independent of the population size $N$ and is bounded by the product of the initial epidemic size $F(0) = g_0$ with the number $1/(1 - R_0)$. Conversely, if $R_0 \geq 1$, then Theorem 2.1 yields that the final epidemic size $g_{\infty}$ will increase with the population size $N$. In general, the population size must be set to be very large (e.g., 10 million) so that the condition for
applying the law of mass action will be justified. Therefore, in the second case ($R_c > 1$), the final epidemic size will be very large and the epidemic is very severe.

The result in Theorem 2.1 (i) gives the baseline for bringing an epidemic course under control: (a) reducing the mean infection rate $r(t)$; (b) shortening the effective mean infectious period $\Sigma(t)$. The goal is to reduce the controlled reproductive number $R_c$ (given by (2.11a)) in such a way that $R_c < 1$. In case a control strategy achieves this, then the ratio $g_{u}/F(t_0)$ of the final epidemic size to the epidemic size at $t_0$ (the intervention time point) is bounded by the constant $1/(1 - R_c)$, which is independent of the population size.

In Table 1 we give some computed values of the solution $p$ of Eq. (2.12c). It is certainly very surprising to see that a disease with a relatively small basic reproductive number $R_0 = 3.0$ ($\gg R_c$) will be able to infect the great part ($\sim 94\%$) of the population if no control measures have been applied.

### Table 1

<table>
<thead>
<tr>
<th>$R_c$</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$ (in %)</td>
<td>18</td>
<td>31</td>
<td>42</td>
<td>51</td>
<td>58</td>
<td>64</td>
<td>69</td>
<td>73</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>$R_c$</td>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
<td>2.4</td>
<td>2.5</td>
<td>2.6</td>
<td>2.7</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>$p$ (in %)</td>
<td>82</td>
<td>84</td>
<td>86</td>
<td>88</td>
<td>89</td>
<td>90</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
</tr>
</tbody>
</table>

3. Eradication and control of diseases: impact of vaccination and public control measures

Which level of herd immunity will be safe? How effective do the control measures have to be, so that an epidemic will be brought under control? These two problems are crucial for the maker of public health policy.

We tackle these two problems by virtue of our new SEIR model Eq. (2.1). Our theory (Theorem 2.1) reveals that the condition for the eradication or control of diseases is that there exists some time point $t_0 > 0$ such that the controlled reproductive number $R_c$ given by

$$R_c := \sup_{t \geq t_0} R_{\text{eff}}(t)$$

is maintained below unity, i.e.,

$$R_c < 1.$$  \hspace{1cm} (3.1a)

$$R_c < 1.$$  \hspace{1cm} (3.1b)

More exactly, we have the following conclusion as a consequence of Theorem 2.1(i): If (3.1b) is satisfied, then an epidemic course will be brought under control after the time point $t_0$ in such a way that the ratio of the final epidemic size to the epidemic size at $t_0$ (of $F(t_0)$ individuals) is bounded by the constant $1/(1 - R_c)$ (which is independent of the population size), no matter how complicated the transmission patterns of the disease are.

In the traditional epidemic theory, it is expected that a disease will be eradicated or an epidemic course will be brought under control if the controlled reproductive number $R_c$ is maintained below unity. However, one important point, namely, the varying patterns of the infection transmission, was ignored by this classical way of thinking. By contrast, in deriving condition Eq. (3.1) we have taken into account all transmission patterns that are most likely to exist and thus condition Eq. (3.1) must be the right one for eradicating the diseases or controlling their outbreaks.

For practical uses, our control condition Eq. (3.1) re-establishes the following baseline of implementing control measures for eradicating or containing a disease: (i) controlling the contact to disease of the public networks. These measures include vaccination, travel restrictions and barrier precautions (e.g., masks, gloves), which lead to reduce the mean infection rate $r(t)$. (ii) Controlling the sources of infection. These measures include vaccination, quarantine, case isolation and contact tracing of infectives, which lead to reduce the effective mean infectious period $\Sigma(t)$.

We take a closer look at the impact of control measures on the latency and recovery processes.

The latent process, i.e., the process that an exposure becomes being infectious, is mainly an intrinsic thing. More exactly, the latent process depends mostly on the disease itself and on the corresponding circumstances like seasons. Certainly, control measures like vaccination can also yield impact on the latent process. However, this impact can be negligible in the practice.

On the contrast, the removed (recovery) process will be strongly affected by the extrinsic control measures. For the purpose of practical applications of our theory, we have to know how the control measures impact the functions $r(t)$ and $A(t, s)$.

Although it is complicated, the assessment of the impact of control measures on infection sources can be measured by the so-called instant efficacy and delayed efficacy, which are defined as follows:

- The instant efficacy $\rho(t)$. At time $t$, we define $\rho(t)$ to be the proportion of infectives that will be blocked by the control measures in such a way that their transmission of infection will be prevented immediately after their emergence and permanently. The efficacies of vaccine, quarantine and case isolation can be put into this category.
- The delayed efficacy $k(t)$ and the delay $\sigma_d$. At time $t$, we define $k(t)$ to be the proportion of infectives that will be blocked by the control measures after its emergence of a mean lifespan of $\sigma_d$ units of time. The efficacy of contact tracing is a typical example of these delay types.

Alternatively, we can define the parameters $(\rho(t), k(t), \sigma_d)$ by decomposing the infectives into three groups:

- **Group I (blocked):** At time $t$, this group of proportion $\rho(t)$ consists of all infectives whose infection is blocked by the control measures immediately and permanently.
- **Group II (blocked after ‘diagnosis’):** At time $t$, this group of proportion $k(t)$ consists of all infectives whose infection will be blocked immediately and permanently after ‘diagnosis’. Here we understand the term ‘diagnosis’ as the ensemble of all control measures such as contact tracing or clinical diagnosis that lead to the detection and isolation of infectives. In this sense, the mean delay $\sigma_d$ can be equally thought of as the waiting time for an infective to be ‘diagnosed’. Let $A_0(t)$ be the recovery CDF for this group. Then we have

$$\int_0^\infty (1 - A_0(t)) \, dt = \sigma_d.$$  \hspace{1cm} (3.2a)

- **Group III (unblocked and ‘undiagnosed’):** At time $t$, this group of proportion $1 - \rho(t) - k(t)$ consists of the remaining infectives whose recovery is intrinsic; that is, this group consists of all infectives for which the control measures do not impact their recovery and thus their recovery process is an intrinsic thing that depends only on the disease itself and on the corresponding circumstances like seasons. The recovery CDF for this group is just the intrinsic CDF $A_0$. The mean value

$$\sigma := \int_0^\infty (1 - A_0(t)) \, dt$$  \hspace{1cm} (3.2b)

is the mean infectious period of the disease. It follows that

$$\sigma_d \leq \sigma.$$  \hspace{1cm} (3.2c)

Under these assumptions, the effective recovery CDF $A(t, s)$ is given by
\[ A(t, s) = \rho(t) + \kappa(t)A_0(s) + (1 - \rho(t) - \kappa(t))A_0(s) \]  
(3.2d)

for all \( s, t \geq 0 \). Correspondingly, the effective mean infectious period \( \Sigma(t) = \int_0^t (1 - A(t, s)) \, ds \) has the form

\[ \Sigma(t) = \kappa(t)\sigma_0 + (1 - \rho(t) - \kappa(t))\sigma. \]  
(3.2e)

To continue, we assume that there is a time point \( t_0 \geq 0 \) after intervention such that

\[ r(t) \leq r_c, \quad \rho(t) \geq \rho_c, \quad \kappa(t) \geq \kappa_c \]  
(3.3a)

for all \( t \geq t_0 \), where \( r_c, \rho_c \) and \( \kappa_c \) are positive constants such that \( \rho_c + \kappa_c < 1 \).

Using condition (3.3a), the effective reproductive number \( R_{eff}(t) \) given by (2.10) can be estimated as follows:

\[ R_{eff}(t) \leq r_c \cdot \Sigma(t) \leq r_c \cdot r_c \sigma_c \quad (\forall t \geq t_0), \]  
(3.3b)

where

\[ \sigma_c := \kappa_c\sigma_0 + (1 - \rho_c - \kappa_c)\sigma \]  
(3.3c)

is called the controlled mean infectious period. It follows that the controlled reproductive number \( R_c \) can be estimated by

\[ R_c = \sup_{t \geq t_0} R_{eff}(t) \leq r_c \sigma_c. \]  
(3.3d)

Below we will treat two more concrete cases: eradication of diseases by mass vaccination, and control of outbreaks by isolation combined with contact tracing and quarantine.

### 3.1. Eradication of diseases by mass vaccination: impact of the vaccine efficacy

We understand the vaccination effect as the ensemble of all impacts of vaccination on the control of diseases that are made before and during an epidemic course.

We consider the situation that in a homogeneously mixed population a proportion \( f \) of the population has been vaccinated after a certain time point \( t_0 \). In order to measure the vaccination effect, we follow Halloran et al. [25] (cf. also [24,41]) by defining two quantities, the vaccine efficacy against infectiousness \( (VEI) \) and the vaccine efficacy against susceptibility \( (VEI) \), as follows:

\[ 1 - VEI := \text{infectiousness of vaccinated infectives} \]

\[ \text{infectiousness of unvaccinated infectives}, \]

and

\[ 1 - VEIS := \text{susceptibility of vaccinated susceptibles} \]

\[ \text{susceptibility of unvaccinated susceptibles}. \]

For simplicity, we assume that the vaccination effect is immediate and permanent. Hence, we can set \( \sigma_0 := 0 \) and \( \kappa_c := 0 \) in (3.3) and thus the effective mean infectious period \( \sigma_c \) given by (3.3c) is

\[ \sigma_c = (1 - \rho_c)\sigma. \]  
(3.4a)

where \( \rho_c \) is the proportion of infectives whose transmission of infection will be blocked by the vaccine immediately and permanently. The quantity \( 1 - \rho_c \) is equal to the fraction of infectives whose transmission of infection has not been blocked by prior immunity and thus can be identified with the effective infectiousness. By setting the infectiousness of unvaccinated infectives to be unity, the value \( 1 - \rho_c \) is calculated by

\[ 1 - \rho_c = f \times (1 - VEI) + (1 - f) \times 1 = 1 - f \times VEI, \]

which yields that

\[ \rho_c = f \times VEI. \]  
(3.4b)

Similarly, the effective susceptibility is equal to \( 1 - f \times VEI \) and thus the controlled mean infection rate \( r_c \) is given by

\[ r_c = (1 - f \times VEI) \times \beta. \]  
(3.4c)

where \( \beta \) is the mean infection rate in the absence of prior immunity.

Remember that the basic reproductive number \( R_0 \) is given by \( R_0 = \beta/(\sigma) \), where \( \sigma \) is the mean infectious period. By (3.4a)–(3.4c), the controlled reproductive number \( R_c \leq r_c \sigma_c \) can be estimated by

\[ R_c \leq (1 - fVEI) \times (1 - fVEI) R_0. \]  
(3.5)

We say that the level \( f \) of herd immunity is safe for the vaccine efficacy of the values \( (VEI, VEI) \) if the controllability condition \( R_c < 1 \) is satisfied. By (3.5), this is the case if

\[ (1 - fVEI) \times (1 - fVEI) R_0 < 1. \]  
(3.6)

It follows that the level \( f \) of herd immunity is safe for the vaccine efficacy \( (VEI, VEI) \) if

\[ f > f_c. \]  
(3.7a)

where \( f_c \), called the critical immunization threshold, is the unique positive solution of the equation

\[ R_0(1 - fVEI)(1 - fVEI) = 1. \]  
(3.7b)

The term ‘critical immunization threshold’ means that the disease can be expected to be eliminated if the herd immunity \( f \) exceeds \( f_c \), no matter how complicated the transmission patterns of the disease vary.

Considering the solution \( f_c \) of Eq. (3.7b) as a function of the basic reproductive number \( R_0 \) and the vaccine efficacy \( (VEI, VEI) \), we see that \( f_c \) increases with \( R_0 \), decreases with \( VEI \) and \( VEI \).

A necessary and sufficient condition for the solution \( f_c \) of Eq. (3.7b) being not greater than 1 is

\[ R_0(1 - VEI)(1 - VEI) < 1. \]  
(3.8a)

which determines a valid domain for the vaccine efficacy \( (VEI, VEI) \). We consider the generic but less unrealistic case that \( VEI = VEI/2 \), i.e., the vaccine efficacy for an infected individual is only half the vaccine efficacy for a non-infected. Then the inequality (3.8a) is satisfied if and only if

\[ VEIS = 2VEIS > \frac{1}{2} \left( 3 - \sqrt{1 + 8/R_0} \right) =: VEIS. \]  
(3.8b)

This implies that the possibility for eradicating the disease by mass vaccination exists if \( VEIS > VEIS \) and \( VEIS > VEIS/2 \). Because of this, we will call \( VEIS \) the generic critical level of vaccine efficacy.

As in the traditional theory of epidemiology [Anderson and May [1]], we consider the somehow idealized case that vaccination works as follows: (i) \( VEIS = 1 \), i.e., a vaccinated individual will have a perfect protection against the disease. (ii) \( VEIS = 0 \), i.e., the vaccination effect falls completely on an infected individual. Under (i) and (ii), we solve Eq. (3.7b) and obtain that

\[ f_c = 1 - 1/R_0 =: q_c \quad (\text{assuming } VEIS = 1, VEIS = 0). \]  
(3.9)

Remember that in the traditional theory of epidemiology [1] the value \( q_c = 1 - 1/R_0 \) is used as the critical immunization threshold for the case that the vaccine efficacy against susceptibility is 100% effective and the vaccine efficacy against infectiousness as well as the varying of transmission patterns can be ignored. Our above theory establishes in a more detailed way the correctness of this traditional theory of epidemiology.

In the following Table 2, we present our estimation for several known diseases.

**Comparison:** The following short list gives the basic reproductive number \( R_0 \) for the known diseases estimated by other different models and methods (in the spirit of fitting the models to concrete data):

- Chickenpox: 7–12 [1]; Diphtheria: 4–8 [1]; Hepatitis B: \( \approx 1.7 \) [60]; Seasonal influenza: \( \approx 1.5 \) [20], 0.9–2.1 (mean: 1.3) [12];
Table 2
Estimation of the basic reproductive number \( R_0 \) and the reproductive number \( R_0^{(s)} \) generated by superspreading events (SSEs)

<table>
<thead>
<tr>
<th>Infection</th>
<th>( t )</th>
<th>( s )</th>
<th>( R_0 ) (type: ( R_0^{(s)} ))</th>
<th>( q ) (SSEs) in %</th>
<th>( VE_s ) (known) in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chickenpox</td>
<td>8–12</td>
<td>10–11</td>
<td>6.3–13.8 (III–IV; 13.8)</td>
<td>81–93 (93)</td>
<td>75–87 (70–90)</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>14–21</td>
<td>2–5</td>
<td>3.5–8.1 (III–IV; 8.1)</td>
<td>71–88 (88)</td>
<td>59–80 (95–97)</td>
</tr>
<tr>
<td>Fifth disease</td>
<td>7–14</td>
<td>4–5</td>
<td>4.6–10.6 (III–IV; 10.6)</td>
<td>78–91 (91)</td>
<td>67–84 (NA)</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>13–17</td>
<td>19–22</td>
<td>2.1–3.4 (II–I; 8.0)</td>
<td>52–71 (88)</td>
<td>40–79 (85–95)</td>
</tr>
<tr>
<td>HFMD(^b)</td>
<td>3–6</td>
<td>3–4</td>
<td>5.4–13.6 (III–IV; 13.6)</td>
<td>81–93 (93)</td>
<td>71–87 (NA)</td>
</tr>
<tr>
<td>Influenza (pan.)</td>
<td>1–3</td>
<td>2–4</td>
<td>1.8–4.4 (I–II; 11.0)</td>
<td>44–77 (91)</td>
<td>33–84 (70–90)</td>
</tr>
<tr>
<td>Measles</td>
<td>6–9</td>
<td>6–9</td>
<td>5.8–14.3 (III–IV; 14.4)</td>
<td>83–93 (93)</td>
<td>73–88 (90–98)</td>
</tr>
<tr>
<td>Mumps</td>
<td>12–18</td>
<td>4–10</td>
<td>4.3–11.2 (III–IV; 11.2)</td>
<td>77–91 (91)</td>
<td>65–85 (90–97)</td>
</tr>
<tr>
<td>Pertussis</td>
<td>7–10</td>
<td>14–19</td>
<td>7.4–18.5 (III–IV; 18.5)</td>
<td>86–95 (95)</td>
<td>78–90 (70–90)</td>
</tr>
<tr>
<td>Polio</td>
<td>1–3</td>
<td>14–20</td>
<td>3.2–7.9 (II–I; 22.0)</td>
<td>69–87 (85)</td>
<td>56–92 (90–100)</td>
</tr>
<tr>
<td>Rubella</td>
<td>7–14</td>
<td>11–12</td>
<td>6.1–15.2 (III–IV; 15.2)</td>
<td>84–93 (93)</td>
<td>74–88 (90–95)</td>
</tr>
<tr>
<td>Scarlet fever</td>
<td>1–2</td>
<td>14–21</td>
<td>3.7–8.0 (II–I; 22.5)</td>
<td>73–88 (96)</td>
<td>61–92 (NA)</td>
</tr>
<tr>
<td>Smallpox</td>
<td>12–14</td>
<td>3–9</td>
<td>4.3–10.8 (III–IV; 10.8)</td>
<td>77–91 (91)</td>
<td>65–84 (100)</td>
</tr>
<tr>
<td>HIV (Aids)</td>
<td>1–15(^c)</td>
<td>1–10(^d)</td>
<td>1.3–6.0 (II–I; 16.2)</td>
<td>23–83 (94)</td>
<td>16–89 (NA)</td>
</tr>
<tr>
<td>SARS(^c)</td>
<td>4–7(^)</td>
<td>7–1(^)</td>
<td>2.0–4.2 (II–I; 10.5)</td>
<td>50–76 (90)</td>
<td>42–87 (NA)</td>
</tr>
<tr>
<td>Ebola</td>
<td>6–8(^e)</td>
<td>3–10(^f)</td>
<td>1.6–3.4 (II–I; 7.9)</td>
<td>38–71 (87)</td>
<td>28–79 (NA)</td>
</tr>
<tr>
<td>AHC(^g)</td>
<td>1–2</td>
<td>3–7</td>
<td>2.2–5.3 (II–I; 13.9)</td>
<td>55–81 (93)</td>
<td>42–87 (NA)</td>
</tr>
<tr>
<td>FMD (farm-level)</td>
<td>2–14</td>
<td>5–8</td>
<td>1.6–4.4 (II–I; 11.0)</td>
<td>38–77 (91)</td>
<td>28–84 (NA)</td>
</tr>
<tr>
<td>Avian influenza</td>
<td>1–3(^g)</td>
<td>10–14(^h)</td>
<td>2.8–6.8 (II–I; 18.8)</td>
<td>64–85 (95)</td>
<td>52–90 (NA)</td>
</tr>
<tr>
<td>Acute HIV-1(^i)</td>
<td>0.5–1.0</td>
<td>1–3.3</td>
<td>7.4–27.7 (III–IV; 27.7)</td>
<td>86–96 (96)</td>
<td>78–93 (NA)</td>
</tr>
<tr>
<td>Acute SIV(^i)</td>
<td>0.5–1.0</td>
<td>1–2.0</td>
<td>6.6–22.0 (III–IV; 22.0)</td>
<td>85–95 (95)</td>
<td>76–92 (NA)</td>
</tr>
</tbody>
</table>

\(^a\) Stands for the Hand, Foot, and Mouth Disease.

\(^b\) The time unit for HIV is year.

\(^c\) Stands for Severe Acute Respiratory Syndrome. Sources for the estimation of \( \tau \) and \( \sigma \) for SARS are: Donnelly et al. [15], Lipsitch et al. [36], [35].

\(^d\) Sources: Breman et al. [4], Khan et al. [34], Okoye et al. [48].

\(^e\) Stands for Acute Hemorrhagic Conjunctivitis. The estimation of \( \tau \) and \( \sigma \) for AHC is taken from CDC [5].

\(^f\) Stands for the Foot-and-Mouth Disease in animals. The estimation of \( \tau \) and \( \sigma \) for FMD covers the known cases, cf. Ferguson et al. [18], Keeling et al. [32], Kao et al. [31], Haydon et al. [26], Chowell et al. [11]. As in [18,32,31,26,11], our model unit is the farm. This is because the spread of FMD between animals within a farm is so rapid that a total farm will be infected immediately if one individual is infected. Therefore, our above estimation for \( R_0 \) is calculated as in [18,32,31,26,11] at the farm level.

\(^g\) According to Stegeman et al. [56], the avian influenza (H7N7) has a mean infectious period of 10–14 days and a mean latent period of about 2 days. Here we assumed that the mean latent period for avian influenza is comparable to the mean latent period for human influenza, ranging from 1 to 3 days.

\(^h\) Our estimation of the mean latent period \( \tau \) and the infectious period \( \sigma \) for HIV-1 cells is based on data given in [49], [50], and [37], by which the latent period \( \tau \) is equal to the reciprocal of the clearance rate \( c \) (i.e., \( \tau = 1/c \)). It was estimated (see Table 1 in [50]) that the values of \( c \) range from 0.5 to 1.0 days, and the values of \( \sigma \) range from 0.51 to 0.96. Based on these considerations, we assume, as given above, that the values of \( \tau \) range from 0.5 to 1.0 days, and \( \sigma \) ranges from 1.0 to 2.0 days.

\(^i\) We have previously used our results on the estimation of the basic reproductive number \( R_0 \) (with or without SSEs (superspreading events)), which are obtained by a method presented in the next section. In Table 2, the types I–IV are defined by the categories to which the diseases belong: Type I–II corresponds to the category ‘mild’ and Type III–IV corresponds to the category ‘severe’.
ler than the basic reproductive numbers generated by the most dangerous diseases like Measles and Pertussis. This might indicate that these diseases do not cause epidemic courses as often as Measles and Pertussis.

The results in Table 2 show that the observed relatively high vaccine efficacy for the diseases Diphtheria, Measles, Mumps, Pertussis, Polio and Rubella should be sufficient for protecting the global spreading of these diseases. Our estimation of the critical level of vaccine efficacy given in Table 2 reveals that a high vaccine efficacy (\(\geq 90\%\)) should be sufficient for eradicating the selected known diseases. This point might be valuable for the medical industry.

- **Control of outbreaks by isolation combined with contact tracing and quarantine**

  The 2002–2003 epidemic SARS provided an example that a newly emerging disease can be contained under the implementation of isolation combined with contact tracing and quarantine; see [53,36,15,58,21,35,2]. Using our models, we can give an explanation for this.

  We decompose the infectives into two groups: (a) Asymptomatic infectives whose transmission occurs prior to symptoms; let \(\theta\) be the proportion of this group. (b) Symptomatic infectives whose transmission occurs just after symptoms. This fraction is equal to \(1 - \theta\).

  The symptom-based public control measures start with a certain time point \(t_0\) (after the outbreak) and work as follows:

  (a) **Isolation**: A proportion of symptomatic infectives are isolated for a long enough duration.

  (b) **Contact tracing and quarantine**: Symptomatic individuals who have been isolated have their contacts traced, and a proportion of their contacts will be quarantined for a long enough duration.

  In order to describe the efficacy of these control measures, we follow the ideas of Fracer et al. [21] and define \(\sigma\) (efficacy of isolation) to be the probability that an infected will be isolated immediately after he/she becomes symptomatic, and \(\eta\) (efficacy of contact tracing) to be the probability that an asymptomatic infective will be detected first by contact tracing and subsequently isolated. For an asymptomatic infective who is detected by contact tracing, generally, there is a time delay between isolation and the starting point of infection. We let \(\sigma_0\) be the mean time of delays of those infectives. Set

\[
\delta := \sigma_0/\sigma \in (0, 1).
\]

(3.10a)

It is reasonable to consider \(\delta\) to be an increasing function of \(\sigma\), the proportion of asymptomatic infectives.

Under above assumptions, the proportion \(\rho_\sigma\) of blocked symptomatic infectives and the proportion \(\kappa_\sigma\) of asymptomatic infectives that are blocked by contact tracing are calculated as follows:

\[
\rho_\sigma = (1 - \theta)\rho_{\eta}, \quad \kappa_\sigma = \theta\kappa_{\eta}.
\]

(3.10b)

Correspondingly, the controlled mean infectious period \(\sigma_\sigma\) given by (3.3c) is calculated by

\[
\sigma_\sigma = [1 - (1 - \theta)\rho_{\eta} - \theta(1 - \delta)\kappa_{\eta}] \cdot \sigma.
\]

(3.10c)

To complete the model settings, we assume, for simplicity, that the scale of the public controls is relatively small and thus the value \(t(t)\), the mean infection rate per unit time per active (unblocked) infective, will be reduced very few, i.e., we assume that \(t(t) = t_\sigma = \beta\), where \(\beta\) is the mean infection rate in the absence of control measures. It follows that the controlled reproductive number \(R_\sigma \leq \tau_\sigma\sigma_\sigma\) can be estimated by

\[
R_\sigma \leq [1 - (1 - \theta)\rho_{\eta} - \theta(1 - \delta)\kappa_{\eta}] \cdot R_0.
\]

(3.11a)

The controllability condition \(R_\sigma < 1\) will be satisfied, if

\[
(1 - \theta)\rho_{\eta} + \theta(1 - \delta)\kappa_{\eta} > 1 - 1/R_0.
\]

(3.11b)

Consequently, we find that an outbreak (with arbitrary types of latent and recovery patterns) will be brought under control by isolation combined with contact tracing and quarantine if these control measures are so effective that the inequality (3.11b) is satisfied. It can be easily seen that (3.11b) is satisfied by the special pair \((\eta, \tau_\sigma) = (1, 0)\) if and only if

\[
\theta < 1/R_0.
\]

(3.11c)

This implies that an outbreak can be brought under control by isolation alone (i.e., \(\eta_1 = 1\) and \(\tau_\sigma = 0\)) if the fraction \(\theta\) of asymptomatic infectives is not greater than \(1/R_0\). Particularly, we find that it is possible to control SARS and relatively mild Smallpox and Influenza by isolation alone, while HIV (\(\theta > 80\%\)) can be controlled only under very effective isolation combined with very effective contact tracing [42]; cf. Table 3.

Similar controllability conditions are obtained by Müller et al. [45] and Fracer et al. [21] using different methods. For example, one result of [21] says that an outbreak can be brought under control by isolation alone, if the proportion \(\eta_1\) of isolated symptomatic individuals is greater than \(1 - 1/R_0/(1 - \theta)\), which is the same result as given by our inequality (3.11b). However, this conclusion of [21] is based on the unrealistic but mathematically simplifying assumption that both latency and recovery processes are exponentially distributed. By contrast, in our derivation of the controllability condition (3.11b) we have taken into account all possible transmission patterns. Thus, our condition (3.11b) must be the right one for controlling an outbreak by isolation combined with contact tracing and quarantine.

We consider the special (but idealized) case that the isolation of symptomatic individuals is perfectly effective and every asymptomatic individual will be detected by contact tracing, i.e.,

\[
\eta_1 = 100\%, \quad \eta_\tau = 100\%.
\]

(3.12a)

Under assumption (3.12a), we find that the controllability condition (3.11b) is satisfied, if

\[
\delta < \delta_\tau := \min\{1, 1/(\theta R_0)\}.
\]

(3.12b)

We call \(\delta_\tau\) the maximally allowed ratio of delay in contact tracing. For controlling a disease, the smaller the value of \(\delta_\tau\) is, the faster the contacts should be traced.

### 4. The estimation of \(R_0\)

As seen in the previous sections, calculating the basic reproductive number \(R_0\) is equivalent to breaking the codes of communicable diseases: One key for a successful eradication program or a control strategy is to know the value \(R_0\) of the disease. In this section, we will present a method which will give a fairly correct estimate of \(R_0\) for most known diseases, and probably more important, for coming unknown diseases.

We prepare some tools. We will use the convolution to simplify the representations. Recall that the convolution \(u * v\) of two

<table>
<thead>
<tr>
<th>Infection</th>
<th>(R_0 (R_0^{\text{std}}))</th>
<th>(\theta(%))</th>
<th>(\theta &lt; 1/R_0)</th>
<th>(\delta) ((\delta_\tau)) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS</td>
<td>2.0–4.2 (10.5)</td>
<td>0–11</td>
<td>Yes</td>
<td>100 (86)</td>
</tr>
<tr>
<td>Influenza (pan.)</td>
<td>1.8–4.4 (11.0)</td>
<td>30–50</td>
<td>If (R_0 &lt; 2)</td>
<td>45–100 (18)</td>
</tr>
<tr>
<td>Smallpox</td>
<td>4.3–10.8 (10.8)</td>
<td>0–20</td>
<td>If (R_0 &lt; 5)</td>
<td>46–100 (46)</td>
</tr>
<tr>
<td>HIV (AIDS)</td>
<td>1.3–6.0 (16.2)</td>
<td>80–100</td>
<td>No</td>
<td>17–96 (6)</td>
</tr>
</tbody>
</table>

The values for \(R_0\) are taken from Table 2, and the values for \(\theta\) are taken from [21] (Fracer et al.). The value \(\delta_\tau\) is calculated by replacing \(R_0\) by \(R_0^{\text{std}}\) and using the maximal value of \(\theta\).
Below we will use this fact without further explanation.

Define the mean infection rate number. Because of this, we postulate that mates of the intrinsic growth rates can yield sometimes drastic transmission patterns and the intrinsic growth rate with which the mean infectious period and all most likely intrinsic growth rates.

K intrinsic growth rate means of our model Eq. (2.1). More exactly, we use Eq. (2.1) to depends on the models used. In our approach, we will do this by calculating the basic reproductive number function.

In particular, the convolution \( 1 \ast u \) is the integral of \( u : 1 \ast u(t) = \int_0^t u(s) \, ds \) (\( t \geq 0 \)).

For a Lebesgue-integrable function \( v \) on \( \mathbb{R} \), satisfying the growth property

\[
sup_{t \geq 0} e^{-\delta t} |v(t)| < \infty \quad \forall \delta > 0,
\]

the Laplace transform \( \mathcal{L}(v) \) is defined by the Lebesgue integral

\[
\mathcal{L}(v) := \int_0^\infty e^{-\lambda t} v(t) \, dt \quad (\text{Re} \, \lambda > 0).
\]

The above definition is extended to a CDF (cumulative distribution function) \( P \) on \( \mathbb{R} \), by defining the Laplace-Stieltjes transform of \( P \), denoted by \( \mathcal{LP} \), through the Stieltjes integral

\[
\mathcal{LP}(\lambda) := \int_0^\infty e^{-\lambda t} dP(t) \quad (\text{Re} \, \lambda > 0).
\]

In particular, if the CDF \( P \) has a density function \( v \), then the Laplace transform of \( v \), coincides with the Laplace-Stieltjes transform of \( P \).

Below we will use this fact without further explanation.

We now turn to our main topic in this section. We note that two ingredients that determine the epidemic spreading are the transmission patterns and the intrinsic growth rate with which the epidemic has started. It was shown by Wearing et al. [59] that different choices of the transmission patterns and different estimates of the intrinsic growth rates can yield sometimes drastically different estimates of the basic reproductive numbers. Because of this, we postulate that the so-called basic reproductive number, \( R_0 \), must be an overall mean, i.e., it is the result of averaging the reproductive numbers generated by all most likely transmission patterns and all most likely intrinsic growth rates.

The formula

\[
R_0 = \beta \sigma
\]

for calculating the basic reproductive number \( R_0 \) as the product of the mean infection rate \( \beta \) and the mean infectious period \( \sigma \) means that the value of \( R_0 \) calculated in this way is a long-time average.

Another baseline for defining \( R_0 \) (Anderson and May [11]) is to define \( R_0 \) as the expected number of secondary infectious cases generated by one typical primary case in an entirely susceptible and sufficiently large population. Since only the primary case is under consideration, the value of \( R_0 \) calculated in this way is completely determined by the dynamic behavior of the epidemic course in the early stages. It is generally known (cf. [11]) that in the early stages the epidemic course grows exponentially with a positive growth rate, \( \lambda \), the so-called intrinsic growth rate of the epidemic. The estimate of the intrinsic growth rate of an epidemic depends on the models used. In our approach, we will do this by means of our model Eq. (2.1). More exactly, we use Eq. (2.1) to model the course of an epidemic spreading in a population that is homogeneously mixed and without prior immunity and control measures. This means that the recovery CDF \( A \) in Eq. (2.1d) is just equal to the intrinsic recovery CDF. In order to save the notation, below we also use \( A \) to denote the intrinsic recovery CDF.

First we need to give a concrete definition of the so-called intrinsic growth rate \( A \) of an epidemic whose course is described by our model Eq. (2.1). For this purpose, we note that in our present case the mean infection rate \( r(t) \) in Eq. (2.1a) is equal to the (constant) intrinsic mean infection rate \( \beta \), i.e., \( r(t) = \beta \). Using \( r(t) = \beta \) and integrating Eq. (2.1a), we obtain

\[
S(t) = S(0)e^{-\beta t/N} = (N - g_0)e^{-\beta t/N},
\]

where

\[
H = \beta(1 - A) * W' * F.
\]

Substituting \( S(t) = (N - g_0)e^{-\beta t/N} \) into Eq. (2.1b) and using \( H = \beta(1 - A) * W' * F \), we obtain

\[
N - F = (N - g_0) \exp(-G + F/N),
\]

where the kernel \( G \) is given by

\[
G := \beta(1 - A) * W'.
\]

By setting \( F_1 := F/N < 1 \) and \( \alpha := g_0/N < 1 \), we rewrite Eq. (4.1) as

\[
\log(1 - F_1) = \log(1 - \alpha) - G + F_1.
\]

In the early stages of the epidemic course, the approximation

\[
- \log(1 - F_1) = F_1 + \sum_{n=1}^{\infty} \frac{1}{n} F_1^n \approx F_1
\]

holds. Therefore, in the early stages of the epidemic course, the solution of the non-linear Eq. (4.2) can be approximated by the solution of the following linear Volterra equation

\[
u = -\log(1 - \alpha) + G * u.
\]

The solution of (4.3) can be expressed as \( u(t) = (-\log(1 - \alpha)) \cdot \psi(t) \) (\( t \geq 0 \)), where \( \psi \) is the solution of the linear Volterra equation

\[
\psi = 1 + G * \psi.
\]

Positivity of the kernel \( G = \beta(1 - A) * W' \) implies that the solution \( \psi \) of (4.4a) satisfies \( \psi(t) \geq 1 \) for all \( t \geq 0 \). Moreover, the exponential growth rate \( A \) of \( \psi \) as defined by

\[
A := \inf \{ \omega \in \mathbb{R} : \exists M > 0 \text{ such that } \psi(t) \leq Me^{\omega t} \text{ for all } t \geq 0 \}
\]

is finite and non-negative. (To see the finiteness of \( A \), we note that \( 0 \leq G < \beta \). Therefore, by choosing a sufficiently large constant \( \omega > 0 \) we have \( \gamma := \int_0^\infty e^{-\gamma t} G(t) \, dt < 1 \). Consider the functions \( v(t) := \psi(t)e^{-\omega t} \) (\( t \geq 0 \)) and \( \alpha(t) := \sup_{s \geq 0} v(s) \). Then a calculation yields that \( v(t) \geq 0 \) satisfies the equation \( v(t) = e^{-\omega t} + \int_0^t (e^{\omega s} G(s)) v(t - s) \, ds \) (\( t \geq 0 \)). It follows that

\[
v(t) \leq 1 + \alpha(t) \int_0^t e^{-\omega s} G(s) \, ds \leq 1 + \gamma \alpha(t)
\]

for all \( t \geq 0 \). Consequently, we have \( \alpha(t) = \sup_{s \geq 0} v(s) \leq 1 + \gamma \alpha(t) \) for all \( t \geq 0 \) and thus \( \alpha(t) \leq 1/(1 - \gamma) = M \). This implies that \( \psi(t) \leq Me^{\omega t} \) for all \( t \geq 0 \), showing the finiteness of \( A \).

We define the value \( A \) given by (4.4b) as the intrinsic growth rate of the epidemic.

Next we derive a relation between \( A \) and the basic reproduction number \( R_0 \). We use the basic theory of linear Volterra equations (cf. [52] Chap. I) to conclude that (4.4a) has a solution with the exponential growth rate \( \lambda > 0 \) if and only if the Laplace transform \( \mathcal{L}(\psi) \) of \( \psi \geq 0 \) is equal to 1 at \( \lambda = A \). As consequence of this, we have

\[
\mathcal{L}(\psi) = 1.
\]

By calculating the Laplace transform of \( G \) we obtain

\[
\mathcal{L}(\psi) = \beta(1 - A)(\lambda)W'(\lambda)
\]

for all \( \lambda \geq 0 \). Thus, by replacing the Laplace transform \( \mathcal{L}(\psi) \) by the Laplace-Stieltjes transform \( \mathcal{L}(\psi) \), we obtain

\[
\mathcal{L}(\psi) = \beta(1 - A)(\lambda)\mathcal{L}(\psi)\lambda
\]

As we will see later, the representation (4.5c) in terms of the Laplace-Stieltjes transform is sometimes more appropriate for our purposes.
Finally, by setting $\lambda = A$ in (4.5c), using condition (4.5a) and replacing $\beta$ by $R_0/\sigma$, we obtain that $R_0 = R_0(A)$, as a function of the intrinsic growth rate $A$, is given by

$$R_0(A) := \frac{\sigma}{(1-A)A} \text{dW}(\lambda)$$  \hspace{1cm} (4.6a)

It can be seen that the function $R_0(\lambda)$ given by (4.6a) is continuously differentiable for all $\lambda > 0$ and satisfies the estimate:

$$R_0(\lambda) \geq \sigma \lambda \quad \forall \lambda > 0.$$ \hspace{1cm} (4.6b)

Moreover, the function $R_0(\lambda)$ is strictly increasing for $\lambda \geq 0$, since the functions

$$(1-A)(\lambda) = \int_0^\infty (1-A(t))e^{-\lambda t} dt.$$ \hspace{1cm} (4.6c)

are strictly decreasing functions of $\lambda > 0$.

To have a real feeling, we compute the function $R_0(\lambda)$ for the special case where both $W$ and $A$ are gamma distributed.

**Example 4.1.** Assume that both $W$ and $A$ are gamma distributed, i.e.,

$$W(t) = \frac{1}{(2\pi)^{1/2}} \frac{1}{(\eta)^{1/2}} e^{-t/(\eta)} A(t) = \frac{1}{(\eta)^{1/2}} (1-e^{-\eta t})$$

with some constants $\eta, \eta > 0$. Then we have that

$$dW(\lambda) = (1+\eta \lambda) \frac{\sigma}{(1-A(\lambda))} (1+\eta \lambda) \frac{1}{(1+\eta \lambda)^{1/2}}.$$ \hspace{1cm} (4.6d)

for all $\lambda > 0$ and thus

$$R_0(\lambda) = \frac{\sigma}{(1-A(\lambda))} (1+\eta \lambda).$$ \hspace{1cm} (4.6e)

for all $\lambda > 0$. In particular, for the case $\eta = 1$ (i.e., where $A$ is exponentially distributed) we have $R_0(\lambda) = (1+\eta \lambda)^2$ which is a strictly convex function of $\lambda > 0$ for every fixed $\lambda \in (0, 1)$. The strict convexity of such functions $R_0(\lambda)$ is important implications, see Proposition 4.3 below.

For the special case $\lambda = 1 = \eta$ (i.e., where both $W$ and $A$ are exponentially distributed) we obtain

$$R_0(\lambda) = (1+\tau_0)(1+\sigma_0) = 1 + (\tau + \sigma) + \tau_0 \sigma^2.$$ \hspace{1cm} (4.6f)

This simple formula was used by Lipschitz et al. [36] to estimate the basic reproductive number for SARS.

- **Average over the intrinsic growth rates:** For the moment, we fix the CDFs $W$ and $A$ that determine the transmission pattern $E_{\lambda=k} = \lambda R$. Theoretically, an epidemic is allowed to start with an arbitrary intrinsic growth rate $A > 0$. However, as observed in the real world, the most likely epidemic courses will start with some typical values of the intrinsic growth rates and thus the basic reproductive number $R_0$ will be given by some typical intrinsic growth rate $A$ in the sense that $R_0 = R_0(A)$.

A crucial problem now is: Which values of the intrinsic growth rates are typical? Or, more concretely, which values of intrinsic growth rates are such that the epidemic will start with them most likely? Mathematically, such typical values of the intrinsic growth rates must obey some kind of abundance.

Below we give a method to determine such typical values of the intrinsic growth rates that will be suitable for calculating the basic reproductive number $R_0$.

We consider an epidemic course with a fixed transmission pattern $E_{\lambda=k} = \lambda R$. An important property, that characterizes also the epidemic course, is the distribution of the secondary cases. Stimulated by the ideas used by Lloyd-Smith et al. in [40], we introduce the ‘individual reproductive number’, $v$, as a random variable representing the expected number of secondary cases generated by a particular infected individual. We assume that $v$ obeys the following distribution law

$$\text{Prob}(v \leq x) = F(x) \quad (x > 0),$$ \hspace{1cm} (4.7a)

where $F : \mathbb{R}_+ \rightarrow [0, 1]$ is some CDF such that $F(0) = 0$ and $F(x) \rightarrow 1$ as $x \rightarrow \infty$. In the sequel, for convenience reasons, we will identify an epidemic course with its distribution law $F$ of secondary cases.

Please do not confuse the notation ‘$F$’ used here with the ones used early in model Eq. (2.1). From now on to the end of this section, the early meaning of ‘$F$’ as the cumulative number of exposures will be dropped. We re-employ the notation ‘$F$’ only for the reason of saving notations.)

In general, the form of $F$ can be arbitrary. We take the mean value

$$F := \int_0^\infty x dF(x)$$ \hspace{1cm} (4.7b)

as the averaged reproductive number generated by an epidemic course that obeys the distribution law $F$.

A relation between the distribution law $F$ and the distribution of the averaged reproductive numbers $R_0(\lambda)$ given by the intrinsic growth rates $A > 0$ can be established as follows. We decompose the group of infected individuals into two subgroups. **Group A** consists of those infected individuals, called typical infected individuals, which have generated at least one secondary case. **Group B** consists of those infected individuals, called atypical infected individuals, which have generated less than one secondary case. B consists of those infected individuals, called atypical infected individuals.

We consider a particular typical infected individual which generates on average $k$ secondary cases ($k \geq 1$). Then there exists a unique $\xi_k > 0$ such that $R_0(\xi_k) = k$. Using our previous formulation by means of model (2.1), it is plausible to think that this typical infected individual generates an epidemic course with the intrinsic growth rate $\xi_k$. Correspondingly, we call $\xi_k$ an individual intrinsic growth rate. In this way, the distribution of the individual intrinsic growth rates $\xi > 0$ is determined as follows:

$$\text{Prob}(\xi > 0 | R_0(\xi) \leq x) = F(x) - F(1)$$ \hspace{1cm} (4.8a)

for all $x > 1$. Since the function $R_0(\xi)$ is strictly increasing for all $\xi > 0$, we find from (4.7a) that the probability distribution of the individual intrinsic growth rates is given by

$$\text{Prob}(\xi > 0 | R_0(\xi) \leq \lambda) = F(R_0(\lambda)) - F(1)$$ \hspace{1cm} (4.8b)

for all $\lambda > 0$. We call the mean value

$$\lambda(F) := \int_0^\infty \lambda d\text{Prob}(\xi > 0 | \xi \leq \lambda) = \int_0^\infty \lambda dF(R_0(\lambda))$$ \hspace{1cm} (4.8c)

the mean intrinsic growth rate of the epidemic course that obeys the distribution law $F$.

The mean intrinsic growth rate $\lambda(F)$ yields the number $R_0(\lambda(F))$ that represents the expected number of secondary cases generated by a particular infected individual which has started its infection chain with the individual intrinsic growth rate $\xi = \lambda(F)$. This local reproductive number $R_0(\lambda(F))$ will coincide with the global reproductive number $F$ if, say, the pre-assigned distribution law $F$ is a point distribution. By a point distribution concentrated at a given point $z > 0$ we mean the CDF $F'$ satisfying

$$dF'(x) = \delta(x - z) dx,$$ \hspace{1cm} (4.9a)
where $\delta(\cdot)$ is the Dirac delta function concentrated at the origin. Equivalently, the function $F$ is given by $F(x) = 0$ for all $0 \leq x < z$ and $F(x) := 1$ for all $x \geq z$. We have for all $A \geq 0$ that

$$\lambda(F_{0}(A)) = A, \quad F_{0}(A) = R_{0}(A). \quad (4.9b)$$

Before going to the next step, we make an observation. Let $A \geq 0$ be a fixed intrinsic growth rate which yields the averaged reproductive number $R_{0}(A)$. By (4.9b) we see that the epidemic course obeying the point distribution law $F_{0}(A)$ has a mean intrinsic growth rate of value $A$ and an averaged reproductive number of value $R_{0}(A)$. We consider a general distribution law $G$ that generates the pair $(\lambda(G), C)$ of mean intrinsic growth rate $\lambda(G)$ and averaged reproductive number $C$. If it occurs that the closedness of the pair $(\lambda(G), C)$ to the given pair $(A, R_{0}(A))$ implies the closedness of the distribution law $G$ to the special point distribution law $F_{0}(A)$, then we can intuitively imagine that there is only one epidemic course, namely the one obeying the special point distribution law $F_{0}(A)$, that will start with the mean intrinsic growth rate $A$ and hit the target $R_{0}(A)$. In other words, the event that an epidemic will start with such a mean intrinsic growth rate $A$ is rare. This observation leads to the following definition.

**Definition 4.2.** We say that an intrinsic growth rate $A \geq 0$ is rare if the corresponding distribution law $F_{A}(A)$ obeys the following absorption property: If $(F_{n})_{n \geq 1}$ is a sequence of CDFs on $\mathbb{R}_{+}$, such that

$$\liminf_{n \to \infty} \lambda(F_{n}) \geq A, \quad \limsup_{n \to \infty} F_{n} \leq R_{0}(A), \quad (4.10)$$

then the sequence $(F_{n})_{n \geq 1}$ converges weakly to the point distribution law $F_{0}(A)$ in the sense that

$$\lim_{n \to \infty} \int_{0}^{\infty} f(x) \, dF_{n}(x) = \int_{0}^{\infty} f(x) \, dF_{0}(A)(x) \quad (4.11)$$

for all continuous and uniformly bounded functions $f$ on $\mathbb{R}_{+}$. An intrinsic growth rate $A$ is not rare in the above sense, then we briefly say that $A$ is typical.

In order to estimate the basic reproductive number of a disease by the averaged reproductive numbers $R_{0}(A)$, only typical intrinsic growth rates $A$ should be taken into account, because the so-called basic reproductive number of a disease is the result of averaging many epidemic courses that might be drastically different.

There is another way to describe rareness. Given an intrinsic growth rate $A \geq 0$, we consider the solutions to the following inequalities:

$$\lambda(F) \geq A, \quad F \leq R_{0}(A), \quad (4.12)$$

where $F$ is a CDF on $\mathbb{R}_{+}$. Clearly, the point distribution $F_{0}(A)$ is a solution of (4.12). Below (in Proposition 4.3) we will see that the rareness of $A$ is equivalent to $F = F_{0}(A)$ being the uniqueness solution of (4.12). This has a very interesting implication. Since for each fixed $A \geq 0$ the set $\mathcal{A}_{A}$ of CDFs that solve (4.12) is closed under convex combinations, we conclude that if $A$ is typical (i.e., not rare), then the solution set $\mathcal{A}_{A}$ is convex and contains infinitely many elements.

A concrete interval containing only typical intrinsic growth rates can be determined as follows. We consider the continuous function

$$h(\lambda) := \lambda/R_{0}(\lambda) = \frac{1}{\sigma} \lambda (1 - A) \lambda \, dW(\lambda) \quad (\lambda \geq 0) \quad (4.13a)$$

that represents the ratio of the intrinsic growth rate to the averaged reproductive number. We have that $h(0) = 0$ and $h(\lambda) \to 0$ as $\lambda \to \infty$. Therefore,

$$M := \sup_{\lambda \geq 0} \lambda/R_{0}(\lambda) \quad (4.13b)$$

is attained. Let

$$\mu := \max\{\lambda > 0 : \lambda/R_{0}(\lambda) = M\}. \quad (4.13c)$$

Furthermore, we have the following results.

**Proposition 4.3.** The following assertions hold.

(i) An intrinsic growth rate $A \geq 0$ is typical if and only if (4.12) admits more than one solution.

(ii) Every $A < \mu$ is typical.

(iii) $\mu$ is rare.

(iv) Assume that the function $R_{0}(\lambda)$ is strictly convex on $\mathbb{R}_{+}$. Then all intrinsic growth rates $A$ with $A \geq \mu$ are rare. Moreover, $\lambda/R_{0}(\lambda)$ attains its absolute maximum at (and only at) $\lambda = A$.

From the above we see that the special intrinsic growth rate $\mu$ obeys the following minimax property: $\mu$ is the smallest number such that every intrinsic growth rate $A < \mu$ is typical, and the largest number among all intrinsic growth rates $A$ such that the ratio $\lambda/R_{0}(\lambda)$ (of the intrinsic growth rate to the averaged reproductive number) attains its absolute maximum at $\lambda = A$.

The proof of Proposition 4.3 will be given in Appendix C. We remark that in most instances the estimator $R_{0}(\lambda)$ given by (4.6a) can be chosen to be strictly convex for all $\lambda \geq 0$. One example was already given in Example 4.1.

Based on the above detailed discussion, we come finally to the following assertion: The basic reproductive number $R_{0}$ is given by some typical intrinsic growth rate $A < \mu$ in the sense that $R_{0} = R_{0}(A)$, i.e.,

$$R_{0} = R_{0}(s) \quad \text{for some} \quad 0 \leq s < 1. \quad (4.14)$$

In general, the scaling factor $s$ is not specified.

- **Average over transmission patterns.** We remember that the function $R_{0}(\lambda)$ given by (4.6a) depends on the pair $(W, A)$ of CDFs that determine the transmission pattern $E \sim I^{+} \sim R$. Recall that both $W$ and $A$ have fixed mean values:

$$\overline{W} = \int_{0}^{\infty} (1 - W(t)) \, dt = \tau, \quad \overline{A} = \int_{0}^{\infty} (1 - A(t)) \, dt = \sigma, \quad (4.15a)$$

where $\tau$ is the fixed mean latent period, $\sigma$ the fixed mean infectious period. Moreover, $W$ obeys the additional condition

$$W(0) = 0. \quad (4.15b)$$

We recall also that the pair $(W, A)$ has the following meanings: (i) $W(t)$ measures how an exposure is far from an infective after his infection of $t$ units of time. (ii) $A(t)$ represents the level of recovery, i.e., $A(t)$ is the probability that an infective will be recovered after having been infective for $t$ units of time.

We try to find some special pairs $(W, A)$ that are ‘typical’, i.e., ‘representatives’ for most of the possible cases. We remark that the ‘representatives’ to be chosen below may not be absolutely continuous (a.c., for short), for two reasons as: (i) These ‘representatives’ have simple forms and with immediate probabilistic interpretations. (ii) These ‘representatives’ describe the cluster of typical a.c. CDFs and thus are limits of such a.c. CDFs.

We characterize a typical transmission pattern $E \sim I^{+} \sim R$ based on the following three properties ($P_{1}$)-(P3):

**P1:** $\overline{W} = \tau$. We assume that $W(t)$ is gamma distributed, that is,

$$W(t) = \frac{1}{(\tau t)^{a}} e^{-t/\tau} \quad (t \geq 0) \quad (4.16a)$$

with some constant $a$, $0 < a < 1$.

**P2:** $\overline{A} = \sigma$. We assume that the recovery probability $A(t)$ grows exponentially, that is,

$$A(t) = 1 - (1 - b) \exp(-(1 - b)t/\sigma) \quad (t \geq 0) \quad (4.16b)$$

with some constant $b$, $0 < b < 1$. 

(P3): The recovery process \( I^R \) changes as \( t \to \infty \) more quickly than its preceding latent process \( E^W \) in the sense that the ratio \( W'(t)/A'(t) \) (of the change of rate) converges to \( t \to \infty \). More precisely, if the pair \((W, A)\) has the forms (4.16a AND 4.16b), then there holds the inequality \( 1/(\alpha \sigma) > (1 - b)/\sigma \), or, equivalently,

\[
\frac{z}{\alpha(1 - b)} > 1 \quad \text{with} \quad z := \frac{\sigma}{\tau}. \tag{4.16c}
\]

We give some comments on these baselines. First, it was demonstrated by P.E. Sartwell [54] that the latent periods of many known diseases are typically 'right-skewed' or 'log-normal' distributed with a long right-hand tail. There are many known 'right-skewed' distributions. However, the gamma distributions are more commonly used and more important, they are more appropriate for our present purpose because their Laplace transforms can be calculated explicitly. Note that in (P1) we have dropped the restriction (4.15b) (saying \( W'(0) = 0 \)) so that the exponential distribution can be adopted. Second, in epidemiology, very often an infected person will show a rash after the latent period. As commonly accepted in the natural sciences, such a rapid change can be modeled by the exponential growth law. Our assumption (P2) follows this basic principle. The assumption (P3) emphasizes the observable fact that the recovery process must, as successor of the latent process, change for large times more quickly than its former.

In order to make the empirical study of the property (P3) more transparent, we rewrite (4.16c) as

\[
\frac{z}{\alpha(1 - b)} = 1 + \frac{z}{\alpha(1 - \eta)} \quad \text{with \ some \} \eta \in [0, 1). \tag{4.16d}
\]

Correspondingly, we define the class \( \{(W_s, A_{s\eta}) : 0 < a \leq 1, 0 < \eta < 1\} \) of CDFs by

\[
W_s(t) := \frac{1}{(\alpha t)\Gamma(1/\alpha)} t^{1 - \alpha} e^{-t^{\alpha}/(\alpha)} \quad (t \geq 0) \tag{4.17a}
\]

and

\[
A_{s\eta}(t) := 1 - \frac{z}{z + z/(1 - \eta)} \exp \left( -\frac{z}{z + z/(1 - \eta)} \cdot \frac{t}{\eta} \right) \quad (t \geq 0). \tag{4.17b}
\]

The Laplace–Stieltjes transforms of \( W_s \) and \( A_{s\eta} \) are

\[
dW_s(\lambda) = \left( 1 + \alpha \tau \lambda^{\alpha} \right)^{-\frac{1}{\alpha}} \frac{\sigma}{\tau} A_{s\eta}(\lambda), \tag{4.17c}
\]

for all \( \lambda \geq 0 \). Remember that the variable

\[
z := \frac{\sigma}{\tau} \tag{4.17d}
\]

used above is the ratio of the mean infectious period to the mean latent period.

We have that

\[
\lim_{\lambda \to 0^+} dW_s(\lambda) = e^{-t^\alpha} \quad (\lambda \to 0) \tag{4.18a}
\]

which is just equal to the Laplace transform \( dW_0(\lambda) \) of the point distribution \( W_0 \) given by

\[
W_0(t) := 0 \quad (0 \leq t < \tau); \quad W_0(t) := 1 \quad (t \geq \tau). \tag{4.18b}
\]

The point distribution \( W_0 \) corresponds to the case of a complete delay of the latent process in the time interval \([0, \tau]\). Such a delay effect has been observed in practice. Hence, the point distribution \( W_0 \) can be accepted as the representative of cases with delay. In short, the ‘representatives’ for the transmission patterns of latency are \([W_s : 0 < s \leq 1]\). Summing up, we have the following assertion: For each pair \((s, \eta) \in [0, 1] \times [0, 1)\) of parameters the transmission pattern \( E^W \) with the choice \((W, A) = (W_s, A_{s\eta})\) satisfies the required properties (P1)-(P3) and thus is a typical transmission pattern.

To continue, we fix a pair \((s, \eta) \in [0, 1] \times [0, 1)\) and denote by \( R_0(s, \eta; \tau; \sigma; \lambda) \) the function \( R_0(\lambda) = \sigma/(1 - A(s, \eta; \lambda)) \) given by (4.6a) with \( W = W_s \) and \( A = A_{s\eta} \). Substituting the Laplace transforms of \( dW_s(\lambda) \) and \( (1 - A_s(\lambda)) \) given by (4.17c), we obtain that

\[
R_0(s, \eta; \tau; \sigma; \lambda) = (1 + \tau \lambda)(1 + \tau \lambda(\pi + z/(1 - \eta))) \quad (\lambda \geq 0). \tag{4.19a}
\]

It is easy to see that the function \( R_0(s, \eta; \tau; \sigma; \lambda) \) is a strictly convex function for all \( \lambda \geq 0 \). Therefore, the value \( \mu > 0 \), which yields the estimate of the basic reproductive number \( R_0 \) by

\[
R_0 = R_0(s, \eta; \tau; \sigma; \mu) \tag{4.19b}
\]

is by Proposition 4.3 (iv) (see also (4.13b and 4.13c)) the unique solution of the equation

\[
d\lambda \frac{\lambda}{R_0(\lambda; \sigma; \tau)} = 0. \tag{4.19c}
\]

Set

\[
x := \lambda \tau, \quad z := \sigma/\tau. \tag{4.20a}
\]

We rewrite the function given by (4.19a) as \( R_0(s, \eta; \tau; \sigma; \lambda) \equiv R_0(\sigma, \eta; x; \lambda) \) with

\[
R_0(\sigma, \eta; x; \lambda) := (1 + 2x)^{-1} [1 + x(\pi + z/(1 - \eta))] \quad (x > 0). \tag{4.20b}
\]

Now a calculation yields that the Eq. (4.19c) becomes a quadratic equation in the new variable \( x = \sigma \mu \) of the form

\[
x + x^2(\pi + z/(1 - \eta)) = 1 + 2x. \tag{4.20c}
\]

Solving (4.20c), we obtain the unique positive solution

\[
x = z/(1 - \sigma + \sqrt{(1 + \sigma)^2 + 4z/(1 - \eta)}). \tag{4.20d}
\]

It follows that the basic reproductive number \( R_0 = R_0(\sigma, \eta; \tau; \sigma; \mu) \) is given by

\[
R_0 = (1 + 2\sigma x)^{-1} [1 + x(\pi + z/(1 - \eta))] \tag{4.21a}
\]

with \( x \) given by (4.20d) and some \( 0 < s < 1 \).

We use the parameters \((s, \eta)\) and the ratio \( z = \sigma/\tau \) to parameterize the scaling factor \( s \), i.e., we take the form \( s = s(\eta, \tau; \sigma) \) as a continuous function of its variables. On the one hand, it is reasonable to assume that \( s(\eta, \tau; \sigma) \to 1 \) as \( z \to \infty \). On the other hand, we need that \( s(\pi; \eta; \sigma) = 0 \). To see this, we note that the solution \( x \) given by (4.20d) has the property that \( x \to 0 \) as \( z \to 0 \). It follows that the function \( R_0 \) given by (4.21a) has the property that \( R_0 \to (1 + 2\sigma (s, \eta; \pi; 0))^{-1} \) as \( z \to 0 \). However, one has to require that \( R_0 \to 1 \) as \( z \to 0 \). Hence, we must have \( s(\pi; \eta; \sigma) = 0 \).

A parameterization \( s = s(\eta; \tau; \sigma) \) that satisfies the above two requirements \((s(\eta, \tau; \sigma) = 0 \) and \( s(\eta, \tau; \sigma) \to 1 \) as \( z \to \infty \)) is, e.g.,

\[
s = \sqrt{4z/(1 - \eta)}/[1 - \sigma + \sqrt{(1 + \sigma)^2 + 4z/(1 - \eta)}]. \tag{4.21b}
\]

We choose such a relatively complicated parameterization form for \( s \) because an \( s \) of the above form, which is closely related to the solution \( x \) given by (4.20d), will yield a very simple representation for \( R_0 \) (see below). We set

\[
y := sy := \frac{4\sqrt{z/(1 - \eta)}}{1 - \sigma + \sqrt{(1 + \sigma)^2 + 4z/(1 - \eta)}} \tag{4.22a}
\]

Let

\[
R_0(s, \tau; z) := (1 + 2x y)^{-1} [1 + y(\pi + z/(1 - \eta))] \tag{4.22b}
\]

with \( y \) given by (4.22a).
The function \( R_0(\alpha, \eta ; z) \) is strictly increasing with \( z \geq 0 \). Moreover, a numerical study reveals that \( R_0(\alpha, \eta ; z) \) is also a strictly increasing function of the parameters \( \alpha \in [0,1] \) and \( \eta \in (0,1) \).

We remember that \( R_0(\alpha, \eta ; z) \) is the basic reproductive number associated with the typical transmission pattern \( E^{\alpha z} I^{\alpha z} R \).

Our final step is to give a method for comparing the different transmission patterns. Our goal for doing this is to derive suitable formulas for estimating the basic reproductive numbers generated by different epidemic courses.

We consider a transmission pattern \( E^{\alpha z} I^{\alpha z} R \) with the parameters \( \alpha \in [0,1] \) and \( \eta \in (0,1) \). Recall that the mode of the gamma distribution \( W_\alpha \) is equal to \((1 - \alpha)\tau \). Clearly, the smaller the mode \((1 - \alpha)\tau \), the faster the latent process \( E^{\alpha z} I \). On the other hand, we note that the growth rate of the recovery probability function \( A_{\alpha R}(t) \) is equal to \( \frac{\alpha}{\alpha + \eta} \). It follows that the smaller the growth rate \( \frac{\alpha}{\alpha + \eta} \), the slower the recovery process \( I^{\alpha z} R \). In short, the transmission pattern \( E^{\alpha z} I^{\alpha z} R \) is characterized by the vector \( \Delta(\alpha, \eta) := (1 - \alpha, z/(\alpha + z/(1 - \alpha))) \).

In the sense that the smaller the vector \( \Delta(\alpha, \eta) \), the faster the latent process \( E^{\alpha z} I \) and the slower the recovery process \( I^{\alpha z} R \). In particular, the larger the parameter \( \alpha \), the faster the latent process \( E^{\alpha z} I \); and the larger the parameter \( \eta \), the slower the recovery process \( I^{\alpha z} R \).

It is an observable fact that a disease with a faster latent process and a slower recovery process will generate more secondary cases than that with a slower latent process and a faster recovery process. Our function \( R_0(\alpha, \eta ; z) \) for estimating the basic reproductive numbers reflects this observed fact very well, since \( R_0(\alpha, \eta ; z) \) is a strictly increasing function of the parameters \( \alpha \in [0,1] \) and \( \eta \in (0,1) \). Therefore, the combinations of the three typical states slow, mean, fast of the latent process \( E^{\alpha z} I \) and the four typical states fast, mean, slow, extremely slow of the recovery process \( I^{\alpha z} R \) yield four typical transmission patterns:

- The pair \( (\alpha, \eta) = (0,0) \) yields the slowest transmission pattern of latency and the fastest transmission pattern of recovery and thus generates the minimal reproductive number \( R_0(0,0 ; z) \).
- We call the transmission patterns \( E^{\alpha z} I^{\alpha z} R \) to be of Type I (the light type).

- The pair \( (\alpha, \eta) = (0.5,0.5) \) yields the mean transmission patterns of latency and recovery and thus generates a mean reproductive number \( R(0.5,0.5 ; z) \). We call the transmission patterns \( E^{\alpha z} I^{\alpha z} R \) to be of Type II (the moderate type).

- The pair \( (\alpha, \eta) = (1,0.95) \) yields the fastest transmission pattern of latency and a slow transmission pattern of recovery and thus generates the maximum reproductive number \( R_0(1,0.95 ; z) \). We call the transmission patterns \( E^{\alpha z} I^{\alpha z} R \) to be of Type III (the severe type).

- The pair \( (\alpha, \eta) = (1,0.99) \) yields the fastest transmission pattern of latency and the extremely slow transmission pattern of recovery and thus generates the largest reproductive number \( R_0(1,0.99 ; z) \). We call the transmission patterns \( E^{\alpha z} I^{\alpha z} R \) to be of Type IV (the extremely severe type).

In the above, we have used the ‘fifty–fifty rule’ to represent the mean processes of latency and recovery. Moreover, we propose to use the ‘95–99% rule’ to determine the slow recovery process: The slow recovery processes are characterized by the CDFs \( A_{\alpha R}(t) \) with \( \eta \) in the range from 0.95 (95%) to 0.99 (99%), i.e., the general slow recovery process corresponds to the choice \( \eta = 0.95 \) and the extremely slow recovery process corresponds to the choice \( \eta = 0.99 \). We remark that there are other possibilities to determine the slow recovery process by choosing other ranges for the parameter \( \eta \). The range \( \eta \in [0.95, 0.99] \) chosen above can be considered as an empirical range, because it yields results which coincide very well with the known estimates, see Table 2 of Section 3 and Table 4 below.

Summing up, we have the following concrete analytical formulas for calculating the basic reproductive numbers.

- \( R_0^0(z) = R_0(1/2, 1/2 ; z) \), the mean reproductive number associated with the transmission patterns of Type II (with the mean latency and recovery processes). By letting \( \alpha := 1/2 \) and \( \beta := 1/2 \) in (4.22a and 4.22b) we have that

\[
R_0^0(z) = (1 + y'(z)z) \exp(y'(z)) \quad \text{with} \quad y'(z) := 2\sqrt{2}/(1 + 2z + \sqrt{1 + 4z}).
\]  
(4.23)

- \( R_0^0(z) \equiv R_0(1.05 ; z) \), the mean reproductive number associated with the transmission patterns of Type II (with the mean latency and recovery processes).
recovery processes). By letting \( \alpha := 1/2 \) and \( \beta := 1/2 \) in (4.22a and 4.22b) we have that
\[
R_0^I(z) = \frac{1}{8} (2 + (1 + 4z)y_I^I(z))^2 \quad \text{with}
\]
\[
y_I^I(z) = 8\sqrt{2z}/(5 + 4z + \sqrt{9 + 8z}). \tag{4.24}
\]
• \( R_{0}^{III}(z) \equiv R_{0}(1.095,z) \), the maximal reproductive number associated with the transmission patterns of Type III (with the fastest latency process and the slow recovery process). Letting \( \alpha := 1 \) and \( \beta := 0.95 \) in (4.22a and 4.22b) we have that
\[
R_{0}^{III}(z) = (1 + \sqrt{20z})(1 + \sqrt{100z}/(1 + 100z)). \tag{4.25}
\]

\* \( R_{0}^{IV}(z) \equiv R_{0}(1.099,z) \), the largest reproductive number associated with the transmission patterns of Type IV (with the fastest latency process and the extremely slow recovery process). Letting \( \alpha := 1 \) and \( \beta := 0.99 \) in (4.22a and 4.22b) we have that
\[
R_{0}^{IV}(z) = (1 + \sqrt{100z})(1 + \sqrt{100z}/(1 + 100z)). \tag{4.26}
\]

Plots of the CDFs for the basic types of transmission patterns are given in Fig. 1. For the plot (b) we have used that the CDF \( B \) of the transmission pattern \( E^- R \), as composition of the latency-recovery transmission pattern \( E^- l^- R \), is given as the convolution of \( A \) with \( dW \), i.e., \( B(t) = \int_0^t \! A(s-t) \, dW(s) \mid t \geq 0 \).

Plots of the reproductive numbers \( R_{0}^{IV}(z) \) as functions of \( z = \sigma/\tau \) are given in Fig. 2. The calculations are performed by virtue of the computer program Mathematica (Wolfram Research). The inequality
\[
1 < R_0^I(z) < R_0^I(z) < R_0^{III}(z) < R_0^{IV}(z) \tag{4.27}
\]
holds for all \( z > 0 \). As seen before, all four functions \( R_0^{IV}(z) \) are strictly increasing functions of \( z = \sigma/\tau \), the ratio of the mean latent period \( \tau \) to the mean infectious period \( \sigma \). This monotonicity of the values of \( R_0^{IV} \) coincides very well with the observed facts. For example, it reveals that acute diseases (i.e., diseases whose latent period is relatively short compared to the infectious period) are more dangerous than chronic diseases (i.e., diseases whose latent period is relatively long compared to the infectious period) in the sense that the former has a larger basic reproductive number than the latter.

**Estimation of \( R_0 \) for most known diseases.** The method for this practice has been explained in the previous section Section 1, and the result has been given in Table 2 (Section 3), in connection with the study of vaccine efficacy and strategies of disease control. In short, we first put a disease, according to the types of its transmission patterns, into one of the two main categories ‘mild’ and ‘severe’ and then estimate its basic reproductive number in terms of the ratio \( z := \text{mean latent period} : \text{mean infectious period} \).

Assuming that a disease of the category ‘mild’ has a mean latent period ranging from \( k \) to \( l \) units of time (e.g., days), and a mean infectious period ranging from \( m \) to \( n \) units of time, then the ratio \( z \) ranges from \( m/l \) to \( n/k \) and thus the basic reproductive numbers \( R_0 \) and the reproductive number \( R_0^{SSEs} \) generated by SSEs for this ‘mild’ disease are estimated by
\[
R_0^I(m/l) < R_0 < R_0^{IV}(n/k), \quad R_0^{SSEs} = R_0^{IV}(n/k). \tag{4.28}
\]

Assuming that a disease of the category ‘severe’ has a mean latent period ranging from \( k \) to \( l \) units of time (e.g., days), and a mean infectious period ranging from \( m \) to \( n \) units of time, then the ratio \( z \) ranges from \( m/l \) to \( n/k \) and thus the basic reproductive numbers \( R_0 \) and the reproductive number \( R_0^{SSEs} \) generated by SSEs for this ‘severe’ disease are estimated by
\[
R_0^I(m/l) \leq R_0 \leq R_0^{IV}(n/k), \quad R_0^{SSEs} = R_0^{IV}(n/k). \tag{4.29}
\]

Our classification for the known diseases is as follows. The category ‘mild’ includes the mild diseases like Hepatitis B, Polio, Scarlet fever and HIV (AIDS) which have a long infectious period (when compared to their mean latent periods) but have an observed small basic reproductive number \( R_0 \). The newly (re-)emerging diseases SARS, Ebola, AHC, FMD, influenza, influenza pandemic and avian influenza are put also into the category ‘mild’, since they have an observed relatively small basic reproductive number \( R_0 \). The category ‘severe’ includes the most severe and extremely severe diseases like Chickenpox, Mumps, Rubella, Measles etc. which have an observed large basic reproductive number \( R_0 \). The newly emerging infections Acute HIV-1 and Acute SIV belong also to the category ‘severe’.

In the following Table 4 we give a more comprehensive overview of our estimates, with a comparison to results obtained by different models (e.g., age-structured models [11]) and methods.
Clearly, the set $X$ is closed and thus an elementary calculation yields that

$$
\frac{\sigma}{\tau} \text{ is the ratio of the mean infectious period to the mean latent period.}
$$

Acknowledgement

The author expresses his deep gratitude to two anonymous referees, to Prof. O. Diekmann, Prof. R. Nagel, Prof. U. Schlotterbeck, Dr. T. Eisner, Ms. B. Dorn and Ms. V. Keicher for their suggestions and help which improve the representations of the manuscript a lot.

Appendix A. Proof of the well-posedness of the SEIR model (2.1)

Integrating the ordinary differential Eq. (2.1a) we obtain that

$$
S(t) = (N - g_0)e^{-H(t)/N}, \quad H(t) := (r(g - R)). \tag{5.1a}
$$

On the other hand,

$$
g(t) - R(t) = \int_0^t (1 - A(s,t) - s)g(s) \, ds \tag{5.1b}
$$

and thus an elementary calculation yields that

$$
H(t) = \int_0^t g(s)K(t - s) \, ds \quad \text{with} \quad K(y, z) := \int_0^y r(s + x)(1 - A(s, x)) \, dx \quad (s, y \geq 0). \tag{5.1c}
$$

By substituting (5.1b) into (5.1a) and using Eq. (2.5) (saying $g = F(0)W + F_1W$ and recalling $F(0) = g_0$) we find finally that our original model Eq. (2.1) are simplified into the following single equation for the function $F$:

$$
F = N - (N - g_0)\exp(-H/N) \quad \text{with} \quad H(t) = \int_0^t K(t - s)(g_0W(s) + F_1\dot{W}(s)) \, ds. \tag{5.2a}
$$

We prove the well-posedness of the SEIR model Eq. (2.1) by showing that the integral (5.2) is well-posed. More exactly, we want to show the following assertion.

**Assertion:** For any given $g_0$, $0 < g_0 < N$, Eq. (5.2) has a unique and global solution $F$ such that $F(0) = g_0$ and $F(t)$ is a non-decreasing and absolutely continuous function of $t \geq 0$ with $F \in L^1\cap (R_+)$.

To this end, we fix $g_0 < 0 < g_0 < N$. For each $t \geq 0$ we set

$$
X_1 := \{u \in AC[0, T] : u(0) = 0, u \leq g_0, 0 \leq u' \leq N \cdot |r|_{L^1(0,T)}\}.
$$

Here and below, we use $AC[0, T]$ to denote the Banach space of all absolutely continuous (real) functions over the interval $[0, T]$. By employing the distance

$$
d(u, v) := \|u' - v'\|_{L^1(0,T)} \quad (\forall u, v \in X_1),
$$

the set $X_1$ becomes a complete metric space.

We consider the set

$$
\Omega := \{T \geq 0 : \exists F \in X_1 \text{ satisfying Eq. (5.2) for all } 0 \leq t \leq T\}.
$$

Clearly, $0 \in \Omega$ and $\Omega$ is closed. We need to show that $\Omega$ is also open in $R_+$. Once the openness of $\Omega$ has been established, we conclude that $\Omega = R_+$ by the connectedness of $R_+$.

We now prove the openness of $\Omega$. It suffices to show that for each $T \in \Omega$ there exists some $\varepsilon > 0$ such that $[T, T + \varepsilon] \subset \Omega$.

For this purpose, fix $T \in \Omega$ and let $u \in X_1$ be the unique solution of Eq. (5.2) in the time interval $[0, T]$. For an arbitrary $M > T$, we define the set

$$
Y_M := \{u \in X_M : u(t) = \rho(t) \quad \forall t \in [0, T]\}.
$$

$Y_M$ is a closed subset of $X_M$ and thus it is also a complete metric space under the metric of $X_M$. Consider the map $P : Y_M \to C[0, M]$ given by

$$
Pu := N - (N - g_0)\exp(-Hu/N) \quad (u \in Y_M), \tag{5.3a}
$$

where the linear map $H : Y_M \to C[0, M]$ is defined by

$$
Hu(t) = \int_0^t K(t - s, s)g_0\dot{W}(s) + u'\dot{W}(s) \, ds \quad (0 \leq t \leq M). \tag{5.3b}
$$

Let $u \in Y_M$. We have

$$
\langle Hu \rangle(t) = \int_0^t r(t)(1 - A(s, t - s))(g_0\dot{W}(s) + u'\dot{W}(s)) \, ds, \tag{5.3c}
$$

which implies

$$
0 \leq \langle Hu \rangle(t) \leq N \cdot |r|_{L^1(0, M)} \quad (t \in [0, M]). \tag{5.3d}
$$

From (5.3c and 5.3d) we see that $Hu \in AC[0, M]$ with $0 \leq \langle Hu \rangle \leq N \cdot |r|_{L^1(0, M)}$. Using this, we find that for each $u \in Y_M$ the function $Pu \in X_M$. Moreover, for each $u \in Y_M$ and $t \in [0, T]$ we have

$$
Pu(t) = N - (N - g_0)\exp(-(1/N)H\rho(t)) = \rho(t) = u(t).
$$

Therefore, $P$ maps $Y_M$ into itself. We want to show that $P$ is a contraction provided that the difference $M - T$ is sufficiently small. To this end, we take $u, v \in X_M$. Then we have $Pu(t) = \rho(t) = Pv(t)$ for all $t \in [0, T]$. For $T \in [T, M]$ we have

$$
Q(t) := Pu(t) - Pv(t) = (N - g_0)(e^{-Hu(t)/N} - e^{-Hv(t)/N}).
$$

It follows that

$$
Q(t) = (1 - g_0/N)[(Hu) - (Hv)](e^{-Hu(t)/N} - e^{-Hv(t)/N})
$$

and thus

$$
(2/(1 - g_0/N))Q(t) = ((Hu) - (Hv))(e^{-Hu(t)/N} - e^{-Hv(t)/N})
$$

+ $(Hu) - (Hv))(e^{-Hu(t)/N} - e^{-Hv(t)/N})$.

This implies by the elementary inequality $|e^y - e^z| \leq |y - z|$ that

$$
\langle Q(t) \rangle \leq \|((Hu) - (Hv))\|_{L^1(0, M)} \cdot \|u' - v'\|_{L^1(0, M)}
$$

On the other hand, we have by (5.3c) that

$$
\|((Hu) - (Hv))\|_{L^1(0, M)} \leq \|u' - v'\|_{L^1(0, M)} \cdot \int_0^T \|W(t)\| \, dt
$$

for $t \in [T, M]$.

Since $u(s) = v(s)$ for all $s \in [0, T]$. It follows that

$$
\int_0^t \|u'(s) - v'(s)\|W(t - s) \, ds \leq \|u' - v'\|_{L^1(0, M)} \int_0^T W(s) \, ds
$$

for all $t \in [T, M]$. Consequently,
\[\| (Hu)^{\prime} - (Hv)^{\prime} \|_{L^\infty[0,M]} \leq \beta_M \cdot d(u,v)\] with \[\beta_M := \int_0^{M-T} W(s) \, ds.\]

This implies, by the inequality \[|Hu(t) - Hv(t)| \leq \int_0^t \| (Hu)^{\prime}(s) - (Hv)^{\prime}(s) \| \, ds,\]
and finding that
\[
\sup_{0 \leq t \leq M} |Hu(t) - Hv(t)| \leq M\beta_M \cdot d(u,v).
\]

Taking these estimates together and using (5.3d) for both functions \([u, v], v \), we find from (5.3e) that
\[\| Q^2 \|_{L^\infty[0,M]} \leq \gamma_M \cdot d(u,v) \] with \[\gamma_M := (1 + 2NM\| F \|_{L^\infty[0,M]} \) \cdot \beta_M.\]

It follows that
\[d(Pu, Pv) \leq \gamma_M \cdot d(u,v) \quad (\forall u, v \in Y_M).\]

Since \[\int_0^{M-T} W(s) \, ds \to 0 \] as \( M \to T \), we conclude that there exists some \( M > T \) such that \( \gamma_M < 1 \). In this case, the mapping \( P : Y_M \to Y_M \) is a contraction. Hence, by the classical contraction principle of Banach we claim that the equation \( Pu = u \) (i.e., Eq. (5.2)) has a unique solution in \( Y_M \). By definition of \( \Omega \), we find that \([T, M] \subset \Omega \). As pointed out before, this implies \( \Omega = \mathbb{R} \), and we thus have proved the well-posedness of Eq. (5.2) as well as the well-posedness of the original SEIR model Eq. (2.1). □

Appendix B. Proof of Theorem 2.1

Before going to the proof of Theorem 2.1, we show first the conclusions in (2.9a)–(2.9c). By the well-posedness proved before the non-decreasing a.c. function \( F(t) \) is bounded by the population size \( N \) and such that \( F \in L^\infty_{\text{loc}}(\mathbb{R}_+) \). Using the monotonicity of \( F \), we have
\[g(t) = \int_0^t F(t-s)W(s) \, ds \leq F(t)W(t) \leq F(t).\]

On the other hand, by the equation \( g'(t) = F(0)W + F' \ast W' (F(0) = g_0) \) again we find that \( g'(t) \geq 0 \) for all \( t \geq 0 \). This implies by Eq. (2.1d) that
\[R(t) = \int_0^t g'(t-s)A(t,s) \, ds \leq \int_0^t g'(t-s) \, ds = g(t),\]

since \( 0 \in A(t,s) \leq 1 \) and \( g(0) = 0 \). This proves Eq. (2.9a).

The bounded and monotone function \( F(t) \) has a finite limit \( g_\infty \) as \( t \to \infty \). We have
\[g(t) = g_\infty \, W(t) + \int_0^t (F(t-s) - g_\infty)W(s) \, ds.\]

It is routine to show that the above integral converges to 0 and thus \( g(t) \to g_\infty \) as \( t \to \infty \). To prove the convergence \( R(t) \to g_\infty \) as \( t \to \infty \), we note that
\[R(t) = g(t) - \int_0^t g'(s) (1 - A(s, t-s)) \, ds.\]

It is routine to show, using the condition \( A(s, t-s) \to 0 \) as \( t \to \infty \) (for each fixed \( s \geq 0 \), since \( A(s, \cdot) \) is a CDF), that the above integral converges to 0 and thus \( R(t) \to g_\infty \) as \( t \to \infty \). To finish the proof of Eq. (2.9b), we use Eqs. (2.1a and 2.1b) to find that
\[F'(t) = -S(t) = \frac{1}{N} r(t)g(t)\, dt, R(t)\, dt.\]

This implies the convergence \( F(t) \to 0 \) as \( t \to \infty \) in (2.9b), since both functions \( r(t) \) and \( S(t) \) are bounded and \( g(t) - R(t) \to 0 \) as \( t \to \infty \). To establish (2.9c), we fix \( t_0 \geq 0 \). We note that
\[\lim_{t \to \infty} K(t-s) = \int_0^\infty g(s+x)(1 - A(s,x)) \, dx = R_{\text{eff}}(s)\]
for each fixed \( s \geq 0 \). Then it follows that
\[\lim_{t \to \infty} H(t) = \lim_{t \to \infty} \int_{t_0}^t K(t-s)g(s) \, ds = \int_0^\infty g(s)R_{\text{eff}}(s) \, ds.\]

This implies, using the representation \( S(t) = S(t_0)e^{-(1/N)[H(t_0)-H(t)])} \), that
\[S(t) = S(t_0)e^{-(1/N)[H(t_0)-H(t)]} + H(t)/N\]
as \( t \to \infty \). Taking the limit for \( t \to \infty \) in (2.1b) (saying \( S(t) + F(t) = N \)) and using the equality \( S(t_0) = N - F(t_0) \), we obtain that
\[(N - F(t_0))e^{-(1/N)[H(t_0)-H(t)]} + H(t)/N\]
which is just (2.9c).

Set
\[x := S(t_0)/N \leq 1, \quad \alpha := F(t_0)/N \leq 1.\]

Then Eq. (5.4a) becomes
\[1 - x = (1 - \alpha)e^{-\eta x}/N \quad \text{with} \quad \eta := \int_0^\infty g(s)R_{\text{eff}}(s) \, ds.\]

We now turn to the proof of Theorem 2.1.

Case (i): \( R_{\text{eff}}(t) \leq R_c < 1 \) for all \( t \geq t_0 \). In this case we have
\[\eta \leq \frac{1}{N} \int_0^\infty g(s)R_{\text{eff}}(s) \, ds \leq \frac{R_c}{N} \int_0^\infty g(s) \, ds = R_c\]

(since \( g(t) \to g_\infty \) as \( t \to \infty \) and thus, by Eqs. (5.4c and 5.4d),
\[1 - x \geq (1 - \alpha)e^{-\eta x} \geq (1 - \alpha)e^{-R_c x}.
\]

Using the elementary inequality \( e^{-R_c x} \geq 1 - R_c x \) \( (x \geq 0) \) we find from the above that
\[1 - x \geq (1 - \alpha)(1 - R_c x),\]

which is equivalent to the inequality
\[\alpha \geq (1 - \alpha)(1 - R_c x).\]

In particular, we have
\[g_\infty /F(t_0) = x /x < 1/(1 - R_c),\]

which is the result (2.11b) of Theorem 2.1 (i).

Case (ii): \( R_{\text{eff}}(t) \geq R_c > 1 \) for all \( t \geq 0 \). In this case we choose \( t_0 = 0 \) and find that
\[\eta \geq \frac{R_c}{N} \int_0^\infty g(s) \, ds = R_c\]

This implies by (5.4c and 5.4d), by noting \( H(t_0) = 0 \) with \( t_0 = 0 \) and \( x > 0 \), that \( x > p \), where \( p \in (0, 1) \) is the unique solution of the equation
\[1 - p = e^{-p \eta}.\]

This is the assertion in Theorem 2.1 (ii).

Case (iii): \( R_{\text{eff}}(t) \geq 1 \) for all \( t \geq 0 \). In this case we choose \( t_0 = 0 \) and find
\[\eta \geq \frac{1}{N} \int_0^\infty g(s) \, ds = 1.
\]

It follows from (5.4c and 5.4d) (with \( H(t_0) = 0 \) for the choice \( t_0 = 0 \)) that
\[1 - x = (1 - \alpha)e^{-\eta x} \geq (1 - \alpha)(1 - x)(1 + x),\]

which yields that \( x \geq \sqrt{2}. \) By the substitution of \( x = g_\infty /N \) and \( x = g_0/N \) we obtain that \( g_\infty \geq \sqrt{g_0 N}. \) This is the assertion in Theorem 2.1 (iii). □
Appendix C. Proof of Proposition 4.3

We first recall the following notations defined in Section 4:
\[ h(\lambda) := \lambda R_0(\lambda) \quad (\lambda \geq 0), \quad M := \max \{ h(\lambda) : \lambda \geq 0 \}, \]
\[ \mu := \max S \quad \text{with} \quad S := \{ \lambda \geq 0 : h(\lambda) = M \}. \]  
Moreover, we let \( \phi(x) (x \geq 1) \) be the inverse function of the strictly increasing function \( R_0(\cdot) \).

Proof of Proposition 4.3

(i): We prove the equivalent assertion stating that \( A \geq 0 \) is rare if and only if (4.12) has the unique solution \( F = F^{R_0(\cdot)} \). Clearly, if \( A > 0 \) is rare, then (4.12) has the unique solution \( F = F^{R_0(\cdot)} \).

Assuming the uniqueness of solutions to (4.12) for a given \( A > 0 \), we want to show that \( A \) is rare in the sense of Definition 4.2.

We consider a sequence \( (F_n)_{n \geq 1} \) of CDFs satisfying (4.10). Let \( (\mu_n)_{n \geq 1} \) be the sequence of the Borel measures induced by \( F_n \). By virtue of the theory of Prokhorov [51], it is easily shown that the uniform bounded condition \( \sup_{n \geq 1} \int_0^\infty x \ dF_n(x) < \infty \) (see (4.10)) implies that the set \( \{ \mu_n : n \geq 1 \} \) is relatively compact in the topology of weak convergence of measures, i.e., every subsequence of \( (F_n)_{n \geq 1} \) contains a weakly convergent subsequence. Hence, in order to show the weak convergence of \( (F_n)_{n \geq 1} \) to \( F \), we may assume that the sequence \( (F_n)_{n \geq 1} \) itself is weakly convergent to some CDF, say \( G \), in the sense that
\[ \lim_{n \to \infty} \int_0^\infty f(x) \ dF_n(x) = \int_0^\infty f(x) \ dG(x) \]  
for all continuous and uniformly bounded functions \( f \) on \( \mathbb{R}_+ \). We want to show that \( G = F^{R_0(\cdot)} \) by proving that \( G \) is a solution of the inequalities (4.12).

Consider the inverse function \( \phi(x) \) of the strictly increasing function \( R_0(\cdot) \). Then we have \( \phi(x)/x \to 0 \) as \( x \to \infty \). It can be shown (details omitted) that the convergence \( \phi(x)/x \to 0 \) as \( x \to \infty \) combined with the boundedness \( \sup_{n \geq 1} \int_0^\infty x \ dF_n(x) < \infty \) implies that (5.6) holds also for the function \( f = \phi \). Hence, we have
\[ \lambda(G) = \lim_{n \to \infty} \int_0^\infty \phi(x) \ dG(x) = \lim_{n \to \infty} \int_0^\infty \phi(x) \ dF_n(x) = \lim_{n \to \infty} \lambda(F_n) \]  
and thus, by condition (4.10),
\[ \lambda(G) = \lambda(F). \]

On the other hand, it is known (e.g., cf. Elstrod [17]) that (5.6) implies the inequality:
\[ \mathcal{G} \int_0^\infty x \ dG(x) \leq \limsup_{n \to \infty} \int_0^\infty x \ dF_n(x) = \limsup_{n \to \infty} \mathcal{F}_n \]  
and thus
\[ \mathcal{G} \leq \mathcal{F} \]  
by condition (4.10) again. Equivalently, the CDF \( G \) is a solution of (4.12). By our uniqueness assumption, we must have \( G = F^{R_0(\cdot)} \). This proves Proposition 4.3 (i).

(ii): We need to show that every \( A < \mu \) is typical. Given \( A < \mu \), it follows from the definition of \( \mu \) that there exists some \( A' > \mu \) such that
\[ A/R_0(A) = A'/R_0(A'). \]  
Consider the distribution law \( H \) given by
\[ dH(x) = (1 - \alpha) \delta(x) dx + \alpha \delta(x - m/\alpha) dx \]  
with \( \alpha := A/A' < 1, \quad m := R_0(A). \)  
Then a calculation yields \( \Pi = R_0(A) \) and \( \lambda(H) \) is such that \( R_0(\lambda(H)/\alpha) = m/\alpha \).

By (5.7a) we have \( m/\alpha = R_0(A') \). Hence, \( R_0(\lambda(H)/\alpha) = R_0(A') \) and thus \( \lambda(H) = \alpha A' = A \), since the function \( R_0(\cdot) \) is strictly increasing. The distribution \( H \) given by Eq. (5.7) is different from the point distribution \( F^{R_0(\cdot)} \) but satisfies the same inequalities (4.12). Therefore, \( A < \mu \) is typical by assertion (i).

(iii): By (i), we need to show the uniqueness of the solution to (4.12) for \( A = 0 \). To this end, let \( G \) be a probability function satisfying (4.12) with \( A = 0 \), i.e., \( \lambda(G) = \mu \) and \( \mathcal{G} \leq R_0(\cdot) \). Then we have, by using the equality \( \mu/R_0(\mu) = M \), that
\[ \int_0^1 M x \ dG(x) + \int_0^\infty x(M - \phi(x)/x) \ dG(x) \leq 0. \]  
where \( \phi \) is the inverse function of \( R_0(\cdot) \). By definition, we have \( \phi(x)/x \leq M \) for all \( x \geq 1 \). Hence, (5.8) implies that the distribution \( G \) is concentrated in the compact set \( [0, \infty) \cup \mathbb{R}(S) \) and thus the composition \( G \circ R_0 \) is concentrated in the compact set \( S = \{ \lambda \geq 0 : \lambda/R_0(\lambda) = M \} \). Since \( \mu > 0 \) is the maximal element in \( S \) and the function \( R_0(\cdot) \) is strictly increasing, we have that
\[ \mu \leq \lambda(G) = \int_0^\infty \lambda \ dG(R_0(\lambda)) \leq \mu \cdot (G(R_0(\mu)) - G(1)). \]  
This implies that the composition \( G \circ R_0 \) is a probability function that is concentrated at the unique point \( x = \mu \). Hence, we must have that \( G = F^{R_0(\mu)} \). This is the desired result.

(iv): Assuming \( R_0(\cdot) \) to be strictly convex on \( \mathbb{R}_+ \), we show that the function \( h(\cdot) = \lambda/R_0(\cdot) \) is strictly increasing for \( 0 < \lambda < \mu \) and strictly decreasing for \( \lambda > \mu \). Clearly, this implies that the set \( S = \{ \lambda \geq 0 : \lambda/R_0(\lambda) = M \} \) is a singleton.

Consider two points \( \lambda_1, \lambda_2 \in [0, \mu] \) with \( \lambda_1 < \lambda_2 \). Let \( x \in (0, 1) \) be such that
\[ \lambda_2 = \lambda_1 + (1 - \alpha) \mu. \]  
Then the strict convexity of \( R_0(\cdot) \) implies that \( R_0(\lambda_2) < \alpha R_0(\lambda_1) + (1 - \alpha) R_0(\mu) \) and thus,
\[ h(\lambda_2) > \frac{\lambda_1}{R_0(\lambda_1)} + \frac{(1 - \alpha) \mu}{R_0(\lambda_1)} \geq \min \left\{ \frac{\lambda_1}{R_0(\lambda_1)}, \frac{\mu}{R_0(\mu)} \right\} = h(\lambda_1), \]  
Since \( h(\lambda_1) = \lambda_1/R_0(\lambda_1) \leq M = \mu/R_0(\mu) \). Thus, we have proved that \( h(\lambda) \) is strictly increasing for \( \lambda < \mu \). The proof of that \( h(\lambda) \) is strictly decreasing for \( \lambda > \mu \) is similar and thus will be omitted. Next we show that every \( A > \mu \) is rare. To this end, let \( A > \mu \) be given and assume that \( G \) is a CDF that solves (4.12), i.e.,
\[ \lambda(G) = \int_0^\infty \lambda \ dG(R_0(\lambda)) \geq A, \quad \mathcal{G} = \int_0^\infty x \ dG(x) \leq R_0(A). \]  
Since \( R_0(\cdot) \) is increasing, convex and \( \int_0^\infty dG(R_0(\lambda)) = 1 - G(1) \), we have by the Jensen inequality that
\[ (1 - G(1)) R_0(A)/(1 - G(1)) \leq \int_0^\infty dG(R_0(\lambda)) \]  
\[ = \int_0^\infty x \ dG(x) \leq R_0(A) - \int_0^1 x \ dG(x). \]  
Equivalently, with \( y := A/(1 - G(1)) \geq A > \mu \) we have that
\[ h(y) \geq A/R_0(A) - \int_0^1 x \ dF(x) \geq A/R_0(A) = h(A). \]  
As shown before, the function \( h(\lambda) \) is strictly increasing for all \( \lambda > \mu \). Hence, we must have \( y = A \), i.e., \( G(1) = 0 \). This yields
which implies by the strict convexity of \( R_0(\cdot) \) that \( G \) must be a point distribution and thus \( G = P\delta^{(1)} \).

**Appendix D. Models with multi-staged exponentially distributed transmission patterns**

For \( z > 0 \) we define the functions \( P_z \) and \( e_z \) by

\[
P^z(t) := 1 - e^{-t/z}, \quad e^z(t) := \frac{1}{z} t e^{-t/z} \quad (t \geq 0).
\]

(5.12a)

The exponential function \( e_z \) is the density function of the CDF \( P_z \). We consider an exponentially distributed transmission pattern \( u \overset{\text{i.i.d.}}{\sim} v \) given by the CDF \( P_z \), i.e.,

\[
v(t) = \int_0^t u(s)e_z(t-s) \, ds \quad (t \geq 0).
\]

(5.12b)

By computing the derivative of \( v \), we obtain that

\[
v'(t) = z^{-1}u(t) - z^{-1} \int_0^t u(s)e_z(t-s) \, ds,
\]

which yields that

\[
zv'(t) + v(t) = u(t).
\]

(5.12c)

In short, the integral (convolution) equation (5.12b) is equivalent to the ordinary differential equation (ode) (5.12c).

Let \( n \) be a positive integer. We call a CDF \( P \) on \( R_+ \) to be \( n \)-staged exponentially distributed if it has the form

\[
P = 1 * e_{\kappa_1} * e_{\kappa_2} * \cdots * e_{\kappa_n}
\]

(5.13a)

with \( n \) positive constants \( \kappa_j > 0 \). Correspondingly, the transmission pattern \( u \overset{\text{i.i.d.}}{\sim} v \) (i.e., \( v = u * P \)) is called \( n \)-staged exponentially distributed if so is the CDF \( P \).

Let \( u \overset{\text{i.i.d.}}{\sim} v \) be an \( n \)-staged exponentially distributed transmission pattern with a CDF \( P \) of the form (5.13a). Then we can decompose the pattern \( u \overset{\text{i.i.d.}}{\sim} v \) into the composition of \( n \) exponentially distributed transmission pattern

\[
u_{0j} := u - u_{1j} - u_{2j} - \cdots - u_{nj} := v
\]

by setting

\[
u_j := u_{j-1} * e_{\kappa_j} \quad (j = 1, \ldots, n).
\]

(5.13b)

As shown before, the system (5.13b) of integral equations can be translated into the following system of ordinary differential equations:

\[
\kappa_j u'_j(t) + u_j(t) = u_{j-1}(t) \quad (j = 1, \ldots, n).
\]

(5.13c)

By defining \( D := \frac{d}{dt} \) to be differentiation, we see that the system (5.13c) can be simplified into one equation

\[
\prod_{j=1}^n (1 + \kappa_j D)u(t) = v(t),
\]

(5.13d)

which is an \( n \)-th order ode for the function \( u \).

A consequence of the above observation is that our model Eqs. (2.1) in Section 2 can be translated equivalently to a system (or delayed system) of odes of the functions \((F, g, R)\), provided that both transmission patterns \( E \overset{\text{i.i.d.}}{\sim} J \) and \( J \overset{\text{i.i.d.}}{\sim} R \) can be decomposed as the sums of multi-staged exponentially distributed patterns and delayed patterns.

Below we compute some special cases. For this purpose, we rewrite Eq. (2.1) in the following more compact form:

\[
F(t) = N - (N - g_0) \exp(-(1/N)H(t)) \quad \text{with}
\]

(5.14a)

\[
H(t) = r(t)g(t) - R(t) - R_0(t) - R_{ad}(t), \quad H(0) = 0,
\]

(5.14b)

\[
g = W' + F \quad \text{and} \quad R(t) = \int_0^t g'(s)A(s,t-s) \, ds.
\]

(5.14c)

We assume first that the latent pattern is 2–staged of the form:

\[
W = 1 * e_a * e_b,
\]

(5.15a)

where \( a, b > 0 \) are positives constants such that

\[
\tau := W = a + b
\]

(5.15b)

is the mean latent period of the disease. Then the convolution equation \( g = W' + F \) in (5.14c) is translated into the following ode:

\[
abg'(t) + \tau g'(t) + g(t) = \tau F(t).
\]

(5.15c)

We assume that the time-dependent CDF \( A(t,s) \) has the form (3.2d) which was used in Section 3 to study the eradication and control of outbreaks, i.e., we take

\[
\frac{A(t,s)}{A(0,s)} = \frac{\rho(t) + \kappa(t)A_0(s) + (1 - \rho(t) - \kappa(t))A_0(s)}{A_0(s)}
\]

(5.15e)

In the above, the function \( A_0 \) is the intrinsic recovery CDF, and \( \rho(t) \) is the time-varying proportion of blocked infectives. Moreover, the function \( \kappa(t) \) is the proportion of infectives that are ‘diagnosed’ and subsequently blocked, and the function \( A_0 \) is the ‘diagnosis’ CDF. The mean value

\[
\sigma_0 = \int_0^\infty (1 - A_0(t)) \, dt
\]

is the mean waiting time that an infective will be ‘diagnosed’. We have that

\[
\sigma_0 \leq \sigma = \int_0^\infty (1 - A_0(t)) \, dt,
\]

where \( \sigma \) is the (intrinsic) mean infectious period of the disease.

Under the choice (5.16a), the function \( R(t) \) given in (5.14c) takes the form

\[
R = R_{\text{inst}} + R_0 + R_{ad}
\]

(5.16b)

with

\[
R_{\text{inst}} = 1 * (\rho g'), \quad R_0 = (\kappa g') * A_0, \quad R_{ad} = ((1 - \rho - \kappa)g') * A_0.
\]

(5.16c)

We assume finally that the intrinsic recovery CDF \( A_0 \) is 2-staged and the ‘diagnosis’ CDF \( A_0 \) is exponentially distributed:

\[
A_0 = 1 * e_a * e_b, \quad A = 1 * e_{\sigma_0},
\]

(5.17a)

where \( c, d, \sigma_0 > 0 \) are positive constants such that

\[
\sigma_0 < c + d = \sigma.
\]

(5.17b)

Under the choice (5.17a), both integral equations in (5.16d) are translated into the following odes:

\[
\sigma_0 R_{0}''(t) + R_{0}'(t) = \kappa(t)g'(t),
\]

(5.18a)

\[
\sigma_0 R_{ad}'(t) + \sigma_0 R_{ad}'(t) + R_{ad}(t) = g(t) - R_{\text{inst}}(t) - (\sigma_0 R_{0}'(t) + R_{0}(t)).
\]

(5.18b)

Summing up, we find that the original system (5.14) with 2-staged exponentially distributed transmission patterns is equivalent to the following second order non-autonomous ode system for the functions \((g, R_{\text{inst}}, R_0, R_{ad})\) :

\[
abg'(t) + \tau g'(t) + g(t) = N - (N - g_0) e^{-(1/N)H(t)},
\]

(5.19a)

\[
H(t) = r(t)g(t) - R_{\text{inst}}(t) - R_0(t) - R_{ad}(t),
\]

(5.19b)

\[
R_{\text{inst}}(t) = \rho(t)g'(t),
\]

(5.19c)

\[
\sigma_0 R_{0}'(t) + R_{0}'(t) = \kappa(t)g'(t),
\]

(5.19d)

\[
\sigma_0 R_{ad}'(t) + \sigma_0 R_{ad}'(t) + R_{ad}(t) = g(t) - R_{\text{inst}}(t) - (\sigma_0 R_{0}'(t) + R_{0}(t)).
\]

(5.19e)
The corresponding initial conditions are:

\[
\begin{align*}
g(0) &= 0 = g'(0), R_{\infty}(0) = 0, R(0) = 0 = R_d(0), \\
R_d(0) &= 0 = R_d(0), H(0) &= 0.
\end{align*}
\] (5.19f)

Remember that all four numbers \(a, b, c, d\) are non-negative, and the sum \(\tau = a + b\) (resp., \(\sigma = c + d\)) is the mean latent (resp., infectious) period of the infection. The number \(N\) is the (fixed) population size, \(g_0 < N\) is the initial epidemic size. Moreover, \(\rho(t) \in [0, 1]\) is the proportion of blocked infectives at \(t\), \(k(t) \in [0,1]\) is the proportion of ‘diagnosed’ infectives at \(t\) and \(\sigma_p\) is the mean ‘diagnosis’ time. Finally, \(R(t) = R_{\infty}(t) + R_0(t) + R_d(t)\) is the cumulative number of removed infectives up to the time point \(t\), and the value of \(R(t)\) can be equally thought of as the number of newly recovered (or newly detected/reported) infectives at time \(t\).

Under the choices (5.16a and 5.17a), the effective mean infectious period \(E(t) = \int_0^\infty (1 - A(t,s))\, ds\) is

\[
E(t) = \kappa(t)\sigma_0 + (1 - \rho(t) - \kappa(t))\sigma. \tag{5.20a}
\]

Assume that there exists a certain time \(t_0 \geq 0\) such that the function \(r(t)\) is decreasing for all \(t \geq t_0\). Then the effective reproductive number \(R_{\infty}(t) = \int_0^\infty r(t+s)\, ds\) can be estimated as follows:

\[
R_{\infty}(t) \leq r(t) \int_0^\infty (1 - A(t,s))\, ds = r(t)\Sigma(t) \quad (t \geq t_0). \tag{5.20b}
\]

If \(\sup_{t \geq t_0} r(t)\Sigma(t) < 1\), then we have

\[
R_\infty(t) = \sup_{t \geq t_0} R_{\infty}(t) \leq \sup_{t \geq t_0} r(t)\Sigma(t) < 1. \tag{5.20c}
\]

It follows from our theory (Theorem 2.1 (i), Section 2) that the epidemic curve described by system (5.19) will be brought under control after time \(t_0\) with a final epidemic size \(g_\infty < F(t_0)/(1 - R)\), where \(F(t_0)\) is the cumulative number of exposures up to time \(t_0\).

In the forthcoming works [29,30] we will use model (5.19) or its variants to study numerically the 2002-2003 epidemic SARS and the vaccination strategy for emergency response to a biological terror attack.

Appendix E. The classical SEIR model as a special case of the new model Eq. (2.1)

We recall (Hethcote [28], Anderson and May [1]) that in the formulation of the classical SEIR model one takes \(S(t), E(t), I(t), R(t)\) to be the number of individuals that are at time \(t\) in the corresponding susceptibility, infectivity, infectiousness and recovery state. As observed in Section 2 (Eq. (2.2b)), both numbers \(E(t)\) and \(I(t)\) can be calculated by our cumulative numbers \(F(t), g(t)\) and \(R(t)\) as follows:

\[
E(t) = F(t) - g(t), \quad I(t) = g(t) - R(t). \tag{5.21}
\]

Moreover, both Eqs. (2.1a and 2.1b) in our model Eq. (2.1) can be rewritten as:

\[
\begin{align*}
S'(t) &= -\frac{r(t)}{N}I(t)S(t), \\
S(t) + E(t) + I(t) + R(t) &= N.
\end{align*} \tag{5.22a, 5.22b}
\]

We consider a special case of our model Eq. (2.1). We assume that both latency and recovery processes are exponentially distributed and the recovery process is time-independent, i.e., both CDFs \(W\) and \(A\) have the following forms

\[
W = 1 + \epsilon_r, \quad A(t, \cdot) \equiv A_0(\cdot) \quad \text{with} \quad A_0 = 1 + \epsilon_r. \tag{5.23}
\]

Under the special choices (5.23), the third equation in our model Eq. (2.1) (Eq. (2.1c)) stating \(g = F \ast W\) becomes the following ode (cf. Eq. (5.15c)):

\[
tg'(t) + g(t) = F(t). \tag{5.24a}
\]

Similarly, the fourth equation in our model Eq. (2.1) (Eq. (2.1d) stating \(R = g \ast A_0 = g \ast A_0(\cdot)\)) becomes the following ode (cf. Eqns. (5.18a and 5.18b)):

\[
sR(t) + R(t) = g(t). \tag{5.24b}
\]

By (5.21), we have \(g(t) = I(t) + R(t)\). Using this in (5.24b) we obtain the following ode for the function \(R(t)\):

\[
R'(t) = \frac{1}{\sigma}I(t). \tag{5.25a}
\]

On the other hand, by using (5.21) again, we have \(g'(t) = I'(t) + R'(t)\) and \(F(t) = g(t) + E(t)\). Substituting both relations into (5.24a) and using (5.25a), we obtain the following ode for the function \(I(t)\):

\[
I'(t) = \frac{1}{\tau}E(t) - \frac{1}{\sigma}I(t). \tag{5.25b}
\]

Finally, by virtue of the identity (5.25b) combined with the odes (5.22a) and (5.25a and 5.25b) we find that the ode for the function \(E(t)\) is

\[
E'(t) = -\frac{r(t)}{N}I(t)S(t) - \frac{1}{\tau}E(t). \tag{5.25c}
\]

Summing up, we have shown that our new model Eq. (2.1) with the special choices (5.23) is equivalent to the following ode system:

\[
\begin{align*}
S'(t) &= -\frac{r(t)}{N}I(t)S(t), \\
E'(t) &= -\frac{r(t)}{N}I(t)S(t) - \frac{1}{\tau}E(t), \\
I'(t) &= \frac{1}{\tau}E(t) - \frac{1}{\sigma}I(t), \\
R'(t) &= \frac{1}{\sigma}I(t),
\end{align*} \tag{5.26a, 5.26b, 5.26c, 5.26d}
\]

System (5.26) is just the classical SEIR model [28,1] with a time-dependent mean infection rate \(r(t)\).

References
