Journal of Theoretical Biology 404 (2016) 295–302

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

The many guises of R_0 (a didactic note)

J.M. Cushing [a,*], Odo Diekmann [b]

[a] Department of Mathematics and Interdisciplinary Program in Applied Mathematics, University of Arizona, 617 N. Santa Rita, Tucson, AZ 85721, USA
[b] Mathematisch Instituut, Universiteit Utrecht, PO Box 80.010, 3508 TA Utrecht, The Netherlands

HIGHLIGHTS

- Derivation of different basic reproduction numbers R_0.
- Construction of the next-generation matrix.
- R_0 and discrete time structured population dynamics.
- R_0 and continuous time models of epidemics.
- R_0 as a determinant of population growth or decay.

ARTICLE INFO

Article history:
Received 7 March 2016
Received in revised form
31 May 2016
Accepted 13 June 2016
Available online 16 June 2016

Keywords:
Basic reproduction number
Next-generation matrix
Population dynamics
Infectious disease dynamics

ABSTRACT

The basic reproduction number R_0 is, by definition, the expected life time number of offspring of a newborn individual. An operationalization entails a specification of what events are considered as "reproduction" and what events are considered as "transitions from one individual-state to another". Thus, an element of choice can creep into the concretization of the definition. The aim of this note is to clearly expose this possibility by way of examples from both population dynamics and infectious disease epidemiology.

1. Introduction

It may be that two population biologists, while dealing with the same model, come up with different numbers, or different expressions in terms of underlying parameters, for the basic reproduction number R_0; see e.g. Bani-Yaghoub et al. (2012). It may happen that both are right. The aim of this short note is to explain the reason for this and to illustrate it with examples.

The key point is that there is sometimes a certain ambiguity in the meaning of "reproduction", in pinpointing what is meant by "newborns" in the bookkeeping framework. Mathematically this is reflected in the fact that there are multiple ways to decompose a positive matrix (in a discrete time model) or positive-off-diagonal matrix (in a continuous time model) into a sum of two matrices with certain properties (as specified in the Appendices). As a consequence, different reproduction numbers, which simply count different things, can result. It is reassuring to know, however, that the sign of $R_0 - 1$ is independent of the decomposition used and that the prediction of exponential growth or decay is therefore correctly made by any of the counting schemes.

As a warm up, we consider in the next section a population of cells that divide into two upon completing the cell cycle. We adopt a generation perspective, meaning that we do not care about (variability in) the length of the cell cycle, but concentrate on the production of offspring. We illustrate how language (in particular, the assigning of name labels) and counting interconnect.

Discrete time models are often used when seasonality is a relentless driver of the life history. In Section 3 we introduce a model of a plant population. The life stages seed-seedling-plant form a cycle, but while both seeds (in the seed bank) and plants (on the field) may a year later still be in the same stage, "seedling" is a one-time-only affair. We first show how this feature enables the straightforward computation of a reproduction number. Next we briefly touch upon the choice of a census point in the year cycle and its influence upon the bookkeeping scheme. We introduce the projection matrix that generates the year-to-year dynamics, i.e. that "projects" the demographic state vector from one census time to the next by means of matrix multiplication (Caswell, 2001, Section 2.5). A decomposition of the projection matrix (satisfying

* Corresponding author.
E-mail addresses: cushing@math.arizona.edu (J.M. Cushing), O.Diekmann@uu.nl (O. Diekmann).

http://dx.doi.org/10.1016/j.jtbi.2016.06.017
0022-5193/

certain conditions delineated in Appendix A) yields a next-generation matrix. R_0 is, by definition, the dominant eigenvalue of the next-generation matrix. Without being exhaustive, we show that several such decompositions are possible and that the corresponding reproduction numbers need not be the same. We argue that the biological interpretation underlying one decomposition may be more natural/convincing than the interpretation of another decomposition. But we stress that this is indeed a matter of interpretation and not of mathematical (in)correctness.

In Section 4 we turn to a continuous time model of the spread of an infectious disease. Following Inaba and Nishiura (2008), we assume that a newly infected individual is immediately infectious and yet does not show any symptoms. After an exponentially distributed amount of time, the asymptomatic individual either loses infectiousness or exhibits symptoms. From the point of view of the infectious agent, transmission is reproduction. On the other hand, the public health system of the human host population labels those developing symptoms as a new case. We elaborate both points of view and explain how the second relates to the control issues that motivated Roberts and Heesterbeek to introduce their type reproduction numbers (Heesterbeek and Roberts, 2007; Roberts and Heesterbeek, 2003); also see Bani-Yaghoub et al. (2012).

For other reflections on R_0 we refer to Heffernan et al. (2005), Heesterbeek (2002), Keeling and Grenfell (2000), Li et al. (2011), and Roberts (2007).

2. A play on words and symbols

Consider a population of single cell organisms, e.g. bacteria. Assume that at the end of the cell cycle, the cell divides into two cells and that each of these immediately starts a cell cycle. Assume that a cell completes the cycle with probability $p \le 1$.

Let us call the cell that divides the "mother" and let us call both cells that arise from the division her "daughters", thus expressing that we consider these as "newborn" and the mother as having died at division. The expected number of offspring of new born individuals is

$$R_0 = 2p \tag{1}$$

since a newborn cell produces with probability p exactly two offspring and with probability $1-p$ no offspring at all.

Alternatively, we might keep using the label "mother" for one of the two cells that exist after the division, while labeling the other of the two as her "daughter". Then a mother produces one offspring at division, but may produce more offspring in the future. The expected number of offspring of a newborn individual is

$$R_0 = p + p^2 + p^3 + \cdots$$

since the newborn cell completes at least n cell cycles with probability p^n and produces one offspring at the end of each completed cycle. If $p = 1$ then the "mother" is, in effect, immortal and produces infinitely many offspring. If we assume

$$p < 1 \tag{2}$$

then the expected number of offspring is finite and

$$R_0 = p(1-p)^{-1}. \tag{3}$$

Let $x(t)$ denote the expected number of cells at time t. Choose as the unit of time the duration of one cell cycle. Then

$$x(t+1) = 2px(t).$$

To motivate a certain matrix decomposition in the next section, we write this as

$$x(t+1) = (T+F)x(t)$$

where T captures survival and F reproduction. So (1) corresponds to the choice $T=0$ and $F = 2p$ while (3) corresponds to the choice $T=p$ and $F=p$. In both cases we have

$$R_0 = F(I-T)^{-1}$$

Since $(I-T)^{-1}$ equals the expected length of life, R_0 is indeed the expected life time offspring production. Note that from a mathematical point of view other decompositions of $2p$ into $T+F$ are perfectly alright.

Clearly the two expressions (1) and (3) differ except when $p=0$ (when both yield $R_0 = 0$) or $p = 1/2$ (when both yield $R_0 = 1$). However, one should not argue about the mathematical correctness of (1) or (3). Either one accepts that there are at least two sensical ways to use the mother/daughter labels, or else one should explain why one way is preferred above the other.

Another simple way to derive (3), when (2) holds, is to perform a *first step analysis*, i.e. to determine R_0 from the consistency equation

$$R_0 = p(1+R_0). \tag{4}$$

That is to say, the expected number of offspring R_0 of a cell is the probability p that it will complete its first cycle, after which it is credited with one offspring with certainty plus an expected number R_0 of future offspring. (This assumes the Markov property that all cells starting the cycle have the same probability to complete it and thus that a mother that survives a cycle is indistinguishable from a newborn cell.)

3. Discrete time models

Consider a plant population with an annual life cycle graph shown in Fig. 1 The arrow from plant to plant corresponds to plants successfully overwintering with probability p_1 while the arrow from plant to seed corresponds to a plant's expected production of f seeds during the summer or fall. A seed germinates and produces a seedling during the spring with probability α_1 which then survives to become a plant with probability α_2 in early summer. An alternative, with probability $1-\alpha_1$, is that a seed does not germinate during the year but instead survives in the seed bank until next year with probability p_2.

Before developing a systematic bookkeeping scheme, let us try to compute a reproduction number directly from its interpretation. The three stages 'plant', 'seed' and 'seedling' constitute a cycle. In

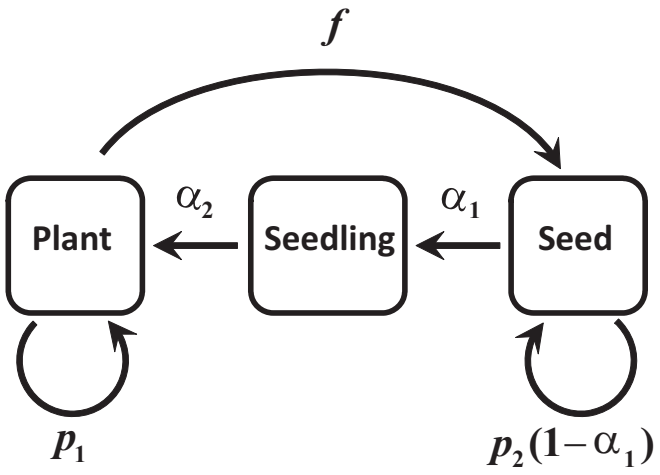

Fig. 1. Plant life cycle.

the life cycle graph there are arrows from seed to seed and from plant to plant, reflecting that both seeds and plants can stay in their stage for several years. In contrast, the stage 'seedling' has only an incoming and an outgoing arrow. So we can consider one seedling and compute relatively easily the expected total number of seedlings that it produces (some after many years). The following algorithm culminates in an explicit formula.

- The expected number of seeds produced by a plant equals (cf. (3))

$$E = f + p_1 f + p_1^2 f + \cdots = f(1-p_1)^{-1}. \tag{5}$$

  Alternatively we derive this formula from the equation

$$E = f + p_1 E$$

  obtained by first step analysis.
- Let Q denote the probability that a newly produced seed develops into a seedling. In Fig. 1, the fact that a seed must survive at least one winter before it can germinate is not clearly expressed. Taking this fact into account in a first step analysis, we obtain the equation

$$Q = p_2\big(\alpha_1 + (1-\alpha_1)Q\big)$$

  and hence

$$Q = \frac{\alpha_1 p_2}{1-(1-\alpha_1)p_2}.$$

- A seedling becomes a plant with probability α_2. As a result R_0, the expected number of seedlings produced by a seedling, is $\alpha_2 EQ$, i.e.

$$R_0 = \frac{f p_2 \alpha_1 \alpha_2}{(1-p_1)\big(1-(1-\alpha_1)p_2\big)}. \tag{6}$$

Although simple and efficient, this computation is ad hoc. A systematic approach requires that we first transform the life cycle graph into a matrix population model (cf. Caswell, 2001; Cushing, 1998)

$$x(t+1) = Px(t). \tag{7}$$

The time variable t in (7) is an integer and counts the years. The population state $x(t)$ is a n-vector and corresponds to a census taken at a particular time in year t. The $n \times n$ matrix P is called the *projection matrix* and its entries should be derived from the life cycle graph in Fig. 1. Both x and P have non-negative elements.

The season in which the census takes place has a subtle influence on the formulation of the model. For instance, if the census occurs in the spring, then we should take $n=3$ (so that the vector x consists of seed, seedling and plant counts), whereas if the census occurs in the summer, autumn or winter, then $n=2$ and the vector x consists of seed and plant counts. As a rule exactly when the census takes place will be determined by the possibilities for gathering data. For the purpose of our exposition, we assume the census takes place in the summer.

Let the first component of x specify the density of plants and the second component specify the density of seeds. Then our assumptions produce the projection matrix

$$P = \begin{pmatrix} p_1 + f p_2 \alpha_1 \alpha_2 & p_2 \alpha_1 \alpha_2 \\ f p_2 (1-\alpha_1) & p_2 (1-\alpha_1) \end{pmatrix}. \tag{8}$$

In order to define a reproduction number, we need to additively decompose the projection matrix

$$P = T + F \tag{9}$$

where F and T are non-negative matrices that, respectively, capture reproduction and population level consequences of mortality and changes in the life stage of individuals (as a mathematical counterpart, inequality (27) should hold).

It makes perfect biological sense to identify reproduction with seed production. We should realize, however, that some seeds produced in autumn show up as plants in the next summer's census. It is helpful to introduce terminology in order to make a distinction between offspring (seeds) that show up as plants in the next census and offspring (seeds) that show up as seeds in the seed bank in the next census. We say that the first have state 1 at birth and the second state 2. Following Rueffler and Metz (2013) we might call "seedling" a hidden state that any seed must go through before becoming a plant that produces seeds. This feature was exploited when we derived (6).

The definition

$$F = \begin{pmatrix} f p_2 \alpha_1 \alpha_2 & 0 \\ f p_2 (1-\alpha_1) & 0 \end{pmatrix} \tag{10}$$

reflects that a seed in the seed bank at census does not produce any offspring in one year, while a plant produces, on average, $f p_2 \alpha_1 \alpha_2$ offspring with state 1 at birth and $f p_2 (1-\alpha_1)$ offspring with state 2 at birth. The transition matrix corresponding to (10), containing the probabilities that plants survive and either remain in the seed bank for another year or become a plant, is

$$T = \begin{pmatrix} p_1 & p_2 \alpha_1 \alpha_2 \\ 0 & p_2 (1-\alpha_1) \end{pmatrix} \tag{11}$$

The matrix F describes expected offspring production in one year. We want to determine the expected *life time* production of offspring. To do so we form, starting from F and T and in the spirit of (3) and the computation of E in (5), the matrix

$$F + FT + FT^2 + \cdots = F(I-T)^{-1} \tag{12}$$

which we call the *next-generation matrix*. Note that the first index of an element in this matrix specifies the state at birth of the offspring, while the second index specifies whether we compute the expected offspring of a plant (index 1) or a seed in the seed bank (index 2).

As motivated and explained in Appendix A, R_0 is defined as the dominant eigenvalue of the next-generation matrix

$$F(I-T)^{-1} = \begin{pmatrix} \frac{f p_2}{1-p_1}\alpha_1\alpha_2 & \frac{p_2\alpha_1\alpha_2}{1-(1-\alpha_1)p_2}\frac{f p_2}{1-p_1}\alpha_1\alpha_2 \\ \frac{f p_2}{1-p_1}(1-\alpha_1) & \frac{p_2\alpha_1\alpha_2}{1-(1-\alpha_1)p_2}\frac{f p_2}{1-p_1}(1-\alpha_1) \end{pmatrix}$$

which is singular (its columns are multiples) and hence has eigenvalue 0. The other eigenvalue is the trace, which is in complete agreement with (6).

The following alternative derivation of (6) from F and T puts more emphasis on the interpretation and saves work when the dimension is higher than two (when computing the inverse of a matrix is troublesome). We have

$$F(I-T)^{-1}v = R_0 v \tag{13}$$

for a nonnegative vector v. The observation that offspring with state 1 at birth and offspring with state 2 at birth are produced in the ratio

$$\alpha_1\alpha_2 : 1-\alpha_1$$

translates into the mathematical statement that the range of F is spanned by the vector

$\begin{pmatrix} \alpha_1\alpha_2 \\ 1-\alpha_1 \end{pmatrix}$.

It follows that any relevant eigenvector v is a multiple of this vector. As a consequence we can compute R_0 by substituting this vector for v in (13). The equation

$$(I-T)y = \begin{pmatrix} \alpha_1\alpha_2 \\ 1-\alpha_1 \end{pmatrix}$$

has solution y with component

$$y_1 = \alpha_1\alpha_2 \frac{1}{\big(1-p_1\big)\big(1-(1-\alpha_1)p_2\big)}.$$

So this substitution results in

$$R_0 = \frac{fp_2\alpha_1\alpha_2}{\big(1-p_1\big)\big(1-(1-\alpha_1)p_2\big)} \tag{14}$$

which is, as noted before, in complete agreement with (6).

In Section 8 of Jin et al. (2015), which is inspired by Eager et al. (2014), a more complicated nonlinear model of a plant population with a seed bank is considered. What follows is not a summary of the analysis in Jin et al. (2015). We expose only a small part of the analysis in order to make our point here. The decomposition of the linearized equation adopted in Jin et al. (2015) and Eager et al. (2014) reduces in our situation to

$$F = \begin{pmatrix} fp_2\alpha_1\alpha_2 & p_2\alpha_1\alpha_2 \\ fp_2(1-\alpha_1) & 0 \end{pmatrix} \tag{15}$$

$$T = \begin{pmatrix} p_1 & 0 \\ 0 & p_2(1-\alpha_1) \end{pmatrix}. \tag{16}$$

The eigenvalues λ of the next-generation matrix

$$F(I-T)^{-1} = \begin{pmatrix} \frac{fp_2}{1-p_1}\alpha_1\alpha_2 & \frac{p_2\alpha_1\alpha_2}{1-(1-\alpha_1)p_2} \\ \frac{fp_2}{1-p_1}\big(1-\alpha_1\big) & 0 \end{pmatrix}$$

are the roots of the quadratic equation

$$\lambda^2 - \frac{fp_2\alpha_1\alpha_2}{1-p_1}\lambda - \frac{fp_2^2\alpha_1\big(1-\alpha_1\big)\alpha_2}{\big(1-p_1\big)\big(1-(1-\alpha_1)p_2\big)} = 0.$$

The positive root of this equation deserves to be called R_0, but generically it is not given by (6).

The reason for the difference in the values of R_0 calculated from the two decompositions (10)–(11) and (15)–(16) resides in the notion of "reproduction", which for (15) is not the same as for (10). Indeed, in (15) the "upgrading" from seed in the seed bank to plant is also considered as a reproduction event, even though the production of the seed in the seed bank was already a reproduction event (an analogy for mammals would be to call both conception and delivery a reproduction event). Admittedly the decomposition (10)–(11) makes more biological sense than the decomposition (15)–(16), but from a mathematical perspective they are equally informative. A slight stretch of the interpretation of the word "reproduction" is all that is needed to interpret R_0 defined by (15)–(16) in the standard manner. Moreover, some algebra shows that the $sign\big(R_0-1\big)$ is the same for both definitions (also see Jin et al. (2015)) and, consequently, both determine whether the population grows or decays. (Indeed, the general theory presented in Appendix A shows that $sign\big(R_0-1\big)$ is the same for any feasible choice of F and T.)

4. Symptomatic versus asymptomatic

We now turn to continuous time and to infectious disease. The considerations below are inspired by Inaba and Nishiura (2008).

We distinguish two kinds of infected hosts, those who are asymptomatic and those who are symptomatic. The first we indicate by index 1, the second by index 2. We concentrate on the initial phase of an epidemic outbreak, meaning that we ignore that the infection process will reduce the availability of susceptible hosts. (Mathematically this amounts to linearization at the disease free steady state. But we shall formulate the linearized problem directly, bypassing the nonlinear problem. This is possible since the linearization at the disease free steady state has an epidemiological interpretation, in contrast to the linearization in an endemic steady state.)

We assume that a newly infected individual is asymptomatic. Asymptomatic individuals become symptomatic with probability η per unit of time. An asymptomatic individual recovers (implying that infectiousness is permanently lost) at rate γ_1 and a symptomatic individual recovers at rate γ_2. An asymptomatic individual produces new infections at rate β_1 and a symptomatic individual does so at rate β_2 (so β_i encodes infectiousness, but the precise value also depends on the density of susceptible hosts).

These assumptions translate into the differential equations

$$\frac{dI_1}{dt} = \beta_1 I_1 + \beta_2 I_2 - \eta I_1 - \gamma_1 I_1$$

$$\frac{dI_2}{dt} = \eta I_1 - \gamma_2 I_2$$

or equivalently

$$\frac{dI}{dt} = AI$$

with

$$A = \begin{pmatrix} \beta_1 - \eta - \gamma_1 & \beta_2 \\ \eta & -\gamma_2 \end{pmatrix}$$

$$I = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}.$$

Here I_1 is the density of asymptomatic individuals and I_2 is the density of symptomatic individuals. In order to define a reproduction number that counts the expected number of secondary cases per primary case, we need to decompose

$$A = T + F$$

where F captures the production of new cases while T captures transitions and removal/recovery.[1] But what is our definition of a "new case"?

The most literal interpretation is to identify "production of a new case" with "transmission of the infectious agent to another host individual". Accordingly we choose

$$F = \begin{pmatrix} \beta_1 & \beta_2 \\ 0 & 0 \end{pmatrix}$$

and

$$T = \begin{pmatrix} -\eta - \gamma_1 & 0 \\ \eta & -\gamma_2 \end{pmatrix}.$$

Note that the range of F is spanned by the vector

[1] In the epidemic literature, Σ is often used to denote transitions and removal/recovery instead of T (Diekmann et al., 2013). A possible source of confusion is that, in that literature, the production of new cases (or transmissions) is denoted by T instead of F.

$\begin{pmatrix}1\\0\end{pmatrix}$,

reflecting that a newly infected individual is asymptomatic.

The components of the (defective) probability vector

$$e^{\tau T}\begin{pmatrix}1\\0\end{pmatrix}$$

describe the chances that an individual is asymptomatic or symptomatic at time τ after infection. Hence the components in the vector

$$\int_0^\infty e^{\tau T}\begin{pmatrix}1\\0\end{pmatrix}d\tau = -T^{-1}\begin{pmatrix}1\\0\end{pmatrix}$$

give the expected amounts of time the newly infected individual will spend asymptomatic and symptomatic respectively. If we multiply the first component by β_1 and the second by β_2, we obtain a reproduction number. Calling this number R_0 we find

$$R_0 = \beta_1\frac{1}{\eta+\gamma_1} + \beta_2\frac{\eta}{\eta+\gamma_1}\frac{1}{\gamma_2}. \tag{17}$$

(Indeed, the newly infected individual remains asymptomatic for an expected time $1/(\eta+\gamma_1)$. With probability $\eta/(\eta+\gamma_1)$ it leaves the asymptomatic phase by developing symptoms. If so, it stays symptomatic for an expected time $1/\gamma_2$.)

From a public health perspective, asymptomatic individuals are "invisible". Taking detectability into account, we might choose to count the production of symptomatic individuals by an individual that just developed symptoms. This corresponds to the choice

$$F = \begin{pmatrix}0 & 0\\ \eta & 0\end{pmatrix} \tag{18}$$

and

$$T = \begin{pmatrix}\beta_1-\eta-\gamma_1 & \beta_2\\ 0 & -\gamma_2\end{pmatrix}. \tag{19}$$

But if $\beta_1 > 0$, then asymptomatic infectives multiply within their own category and one wonders whether this can work. Indeed, if

$$\frac{\beta_1}{\eta+\gamma_1} > 1$$

then an asymptomatic infective produces, on average, more than one new asymptomatic infective before it either develops symptoms or recovers. So, in this case, we get exponential growth of asymptomatic infectives even if $\beta_2 = 0$, i.e. even if the contribution of symptomatic individuals to transmission is ignored. To avoid this, we require

$$\beta_1 < \eta+\gamma_1.$$

This condition also guarantees that T is invertible and that

$$-T^{-1} = \frac{1}{(\eta+\gamma_1-\beta_1)\gamma_2}\begin{pmatrix}\gamma_2 & \beta_2\\ 0 & \eta+\gamma_1-\beta_1\end{pmatrix}$$

is a non-negative matrix. After noting that the range of F defined by (18) is spanned by

$$\begin{pmatrix}0\\1\end{pmatrix}$$

we define R_0 by

$$-FT^{-1}\begin{pmatrix}0\\1\end{pmatrix} = R_0\begin{pmatrix}0\\1\end{pmatrix}$$

from which we compute

$$R_0 = \beta_2\frac{\eta}{(\eta+\gamma_1-\beta_1)\gamma_2}. \tag{20}$$

In order to illuminate the interpretation of (20), we reformulate its derivation. An individual that just developed symptoms remains infectious for an expected amount of time $1/\gamma_2$ and hence produces on average β_2/γ_2 asymptomatic individuals. Let E denote the expected number of individuals that enter the symptomatic class while being a "descendant" of a newly produced asymptomatic individual. Then $R_0 = \beta_2 E/\gamma_2$. It remains to calculate E.

With probability $\eta/(\eta+\gamma_1+\beta_1)$ a newly produced asymptomatic individual will develop symptoms before either recovering or producing, by transmission, another asymptomatic individual. Similarly, the probability that the first event is recovery equals $\gamma_1/(\eta+\gamma_1+\beta_1)$ (in which case there result no symptomatic individuals at all), while the probability that the first event is transmission equals $\beta_1/(\eta+\gamma_1+\beta_1)$ (in which case the expected number of descendants becomes $2E$). Hence we have

$$E = \frac{\eta}{\eta+\gamma_1+\beta_1}\cdot 1 + \frac{\gamma_1}{\eta+\gamma_1+\beta_1}\cdot 0 + \frac{\beta_1}{\eta+\gamma_1+\beta_1}2E$$

from which we conclude that

$$E = \frac{\eta}{\eta+\gamma_1-\beta_1}.$$

The issue of detectability (our motivation for the choice (18)–(19)) is closely related to the issue of control. By targeted control efforts we might be able to reduce β_2 and/or increase γ_2. Now recall that $R_0 = \beta_2 E/\gamma_2$ and note that E is expressed in terms of parameters pertaining to asymptomatic infectives. So for R_0 defined by (20), control efforts targeted on symptomatic infectives have a multiplicative effect, whereas for R_0 defined by (17) such efforts have a multiplicative effect on one term only. It is exactly this difference that motivated Roberts and Heesterbeek (2003) (also see Heesterbeek and Roberts (2007) and Bani-Yaghoub et al. (2012)) to introduce the type-reproduction number. We refer again to Inaba and Nishiura (2008) for a detailed elaboration in the context of a far more general model involving the distinction between symptomatic and asymptomatic infectives.

According to the general theory presented in Appendix B we have sign $(R_0 - 1) = \text{sign}(r)$, independently of the decomposition used to define R_0, where r is the spectral bound of the matrix A or, in more biological jargon, the Malthusian parameter (aka the intrinsic rate of natural increase). In particular, our conclusion about asymptotic stability ($R_0 < 1$) or instability ($R_0 > 1$) does not depend on the decomposition used to define R_0.

5. Concluding remarks

As noted in the abstract, R_0 is defined as the expected life time number of offspring of a newborn individual. In the appendices, however, we define R_0 as the spectral radius of a next-generation matrix K. Specifically, in the discrete time population dynamic setting, $K = F(I-T)^{-1}$ and in the continuous time infectious disease setting $K = -FT^{-1}$. Each is a non-negative matrix and therefore its spectral radius is a dominant positive eigenvalue which has a nonnegative eigenvector y (Berman and Plemmons, 1994, Theorem 2.1.1):

$$Ky = R_0 y. \tag{21}$$

(See Diekmann et al. (2010) for a meaningful way to reduce the dimension of K that eliminates zero elements of y.) How does the definition of R_0 given in the abstract relate to this more technical characterization?

In the word "expected" we need to include the distribution of

type at birth in the following way. Assume the eigenvector y in (21) is normalized such that its components y_i sum to one. Then y_i can be interpreted as the probability that a newborn individual is of type i. Interpreted this way, the statement (21) indeed amounts to the statement: a newborn individual produces, on average, R_0 offspring. But what if we start with a population that does *not* have a distribution of type at birth as described by y? As a rule, R_0 is a *strictly* dominant eigenvalue. As a result, if we iterate K starting from an arbitrary distribution of types at birth, the resulting vectors become more and more like a multiple of y (i.e., the relative size of the components is given by y). In other words, the effects of the particular way in which the population (or the infectious agent) was introduced die out. So we just need a bit of patience. (Yet we shouldn't be too patient, as nonlinear effects gain importance when the population grows. We refer to page 175 of Diekmann et al. (2013) for some more discussion of this aspect.) As far as we know little can be said in general about the exceptional case that R_0 is not strictly dominant.

The key message of this short note is that there can be multiple ways to associate a next-generation matrix with a given matrix that generates the real time dynamics. Accordingly, there are multiple reproduction numbers that deserve to be denoted by R_0, however confusing that may be. To tell them apart, one has to pay attention to the decomposition (of the projection matrix or its continuous time analog) that underlies the next-generation matrix. This decomposition "defines" those events one considers as reproduction events and hence what exactly is being counted. Everybody should feel free to argue that one decomposition is more biologically meaningful than another. Or that one is more relevant for determining the required control efforts than another. Such arguments may in fact be illuminating and/or helpful, but one cannot argue that one is, according to mathematical theory, the one and only right one.

The paper (Browne and Webb, 2015) (brought to our attention by Horst Thieme) is a case in point. If, apart from the infection status of patients, the contamination status of hospital rooms is also incorporated, a complicated transmission model results. The matrix resulting from linearization at the disease free state allows multiple meaningful decompositions (the paper mentions two, but one can easily come up with arguments leading to yet another choice). If, as is indeed the case in this paper, the ultimate aim of the model is to provide insight for control, one may in fact start by listing the parameters that one hopes to be able to control and base the choice of F on these. The motivation is provided by the formulas (30) and (31) in Appendices A and B, which show the scalar multiplicative reduction of F needed to achieve eradication. This observation simply repeats the main idea of Heesterbeek and Roberts (2007); Inaba and Nishiura (2008); Roberts and Heesterbeek (2003), but puts it in a wider and, we think, useful context.

Acknowledgment

We thank KaYin Leung for helpful comments on a draft version of the paper and for bringing (Bani-Yaghoub et al., 2012) to our attention. We also thank Horst Thieme and an anonymous referee for constructive suggestions that helped improve the manuscript.

Appendix A. Discrete time models

Matrix iterations of the form

$$x(t+1) = Px(t). \tag{22}$$

describe the discrete time dynamics of populations structured by a finite number of states at the individual level. The population at time t is described by the n-vector $x(t)$ of state densities and P is an $n \times n$ matrix, called the *projection matrix*. Both x and P have non-negative entries.

paper Li and Schneider (2002) provide an excellent concise introduction to the relevant notions and results concerning the reproductive number for matrix population models. (For more detailed accounts see Caswell (2001) and Cushing (1998).)

For any matrix M we denote the *spectral radius* of M by $\rho(M)$ and recall that one can define

$$\rho(M) = \max\{|\lambda|:\ \lambda \text{ is an eigenvalue of } M\} \tag{23}$$

and show that (where k is an integer)

$$\rho\left(M\right) = \inf_{k\geq 1} \left\|M^k\right\|^{1/k} = \lim_{k\to+\infty} \left\|M^k\right\|^{1/k} \tag{24}$$

holds, or conversely define $\rho(M)$ by (24) and show that (23) holds. A key point is that whenever $\rho(M) < 1$ then $I - M$ is invertible and

$$(I-M)^{-1} = I + M + M^2 + \cdots. \tag{25}$$

In Li and Schneider (2002) (also see Cushing and Yicang (1994)) the projection matrix is additively decomposed

$$P = T + F \tag{26}$$

where F and T are non-negative matrices that, respectively, capture reproduction and population level consequences of changes in the state of individuals. In analogy to (2) it is required that

$$\rho(T) < 1, \tag{27}$$

which excludes immortality in the sense that it guarantees that $\lim_{k\to\infty} T^k x = 0$ for all nonnegative vectors x (k is an integer) (Li and Schneider, 2002). Usually the column sums of T do not exceed one, with at least one column sum less than one, which implies (27). The upshot is that (25) holds for $M=T$ and

$$F(I-T)^{-1}$$

yields a matrix analog of the right side of (3).

We assume the projection matrix P is irreducible. This means there is a path between any pair of population states by means of reproduction and transitions. Mathematically, it means that no reordering of the states will bring P into upper triangular block form. Perron–Frobenius theory (Berman and Plemmons, 1994) guarantees that the *population growth rate*

$$r = \rho(P)$$

is a positive, simple, and dominant eigenvalue of P (in the sense that $|\lambda| \leq r$ for any eigenvalue λ of P) with associated positive right and left eigenvectors. If $r < 1$ the population decays exponentially, if $r > 1$ the population grows exponentially, and if $r=1$ the population is stable. It turns out that the *per generation growth rate* (sometimes called the *net reproduction number*)

$$R_0 = \rho\left(F(I-T)^{-1}\right) \tag{28}$$

and r are on the same side of 1 (Cushing and Yicang, 1994) or more precisely one of the following holds:

$$0 \leq R_0 \leq r < 1 \text{or } 1 < r \leq R_0 \text{or } r = 1 = R_0 \tag{29}$$

(Theorem 3.3 in Li and Schneider (2002)). *It follows that both r and R_0 can be used to determine the growth or decay of the population.* One important fact is that formulas are sometimes more readily available for R_0 than for r. This is because the number of birth states is often low (often equal to one) which means that F is low rank. In addition we have, provided $R_0 > 0$,

$$\rho\left(\frac{F}{R_0}+T\right)=1 \tag{30}$$

which can be interpreted as saying "in order to stop population growth one has to reduce reproduction by a factor $1/R_0$" (Li and Schneider, 2002, Theorem 3.1). Another often useful fact is that

$$\rho\left(F(I-T)^{-1}\right)=\rho\left((I-T)^{-1}F\right).$$

This means the calculation of $(I-T)^{-1}$ need be made on only the range of F, which is what we used to calculate (14).

Appendix B. Continuous time models

The *spectral bound* $s(M)$ of a matrix M is defined by

$$s(M)=\sup\{\mathrm{Re}\,\lambda:\ \lambda \text{ is an eigenvalue of } M\}.$$

Suppose the off-diagonal entries of M are non-negative. Then

$s(M)<0$ if and only if M is invertible and $-M^{-1}$ is non–negative.

One way to understand this is to observe that

$$\int_0^\infty e^{\tau M}d\tau=-M^{-1}$$

if the integral converges. (Also, one can apply Theorem 2.3, parts G_{20} and N_{38} in Berman and Plemmons (1994) to $-M$, which is a so-called M-matrix.)

Assume the coefficient matrix A in the linear ODE system

$$\frac{dx}{dt}=Ax$$

has non-negative off-diagonal entries. Then the system preserves non-negativity (i.e. $\exp(At)\geq 0$ for all $t\geq 0$, Theorem 3.12 in Berman and Plemmons (1994)). Let the decomposition

$$A=T+F$$

be such that T has non-negative off-diagonal entries and F is non-negative. Assume that $s(T)<0$ (in order to exclude immortality). Define

$$R_0:=\rho\left(-FT^{-1}\right)$$

and

$$r=s(A).$$

Then

$$sign\,(r)=sign(R_0-1).$$

See Thieme (2009) and Diekmann et al. (2010), Theorem A1, but be aware of some notational differences: T there corresponds to F here, while T here is called Σ there, and "positive" there corresponds to "non-negative" here.

In addition we have, provided $R_0>0$,

$$s\left(\frac{F}{R_0}+T\right)=0 \tag{31}$$

which can be interpreted as saying "in order to stop population growth one has to reduce reproduction by a factor $1/R_0$".

Appendix C. Miscellanea

The aim of this final appendix is to provide some pointers to the literature concerning aspects of R_0 that are not directly related to the matter of choice in the $T+F$ decomposition.

The generation bookkeeping presupposes that it does not matter when offspring is produced. If, however, the environmental conditions vary in the course of time, it does matter when an individual is born. So does the concept of R_0 perish when the environment is not constant? In Section 7.9 of Diekmann et al. (2013) it is explained how a simple trick can save us when the environmental conditions are periodic. The idea is to label newborn individuals with the phase φ in the cycle at the time of their birth. The next-generation operator then maps functions of φ to functions of φ, so acts on an infinite-dimensional space. That makes it harder to compute R_0, but conceptually nothing changes. See Bacaër and Ait Dads (2012), Cushing and Ackleh (2012), and Wang and Zhao (2008).

There are other contexts in which the individual state space is infinite dimensional. A prominent example arises when we distinguish individuals according to their geographical position. Then it may easily happen that the operator T is unbounded, for instance when T is the Laplace operator describing diffusion. Yet $-T^{-1}$ (or in the discrete time case $(I-T)^{-1}$) may be bounded so that R_0 is still the dominant eigenvalue of a bounded positive operator. The key reference is Thieme (2009).

When derived by using the $F+T$ decomposition, the next-generation matrix (or operator) naturally is the product of two matrices (or operators). If we consider the transpose of the matrix (the adjoint of the operator) the order of the two factors reverses. The dominant eigenvalue does not change in the process, but the corresponding eigenvector does change. While the eigenvector of the next generation matrix provides (when suitably normalized) the stable distribution of birth states, the adjoint eigenvector yields Fisher's reproductive value (Caswell, 2001). So both have meaning, but the meaning differs. In the periodic setting, the order of the two factors gave rise to a little controversy about the 'correct' definition of R_0, see the references in Diekmann et al. (2013), Section 7.9.

The linear next-generation operator ignores that it might take two to reproduce. For a mechanistic derivation of an alternative we refer to Heesterbeek and Metz (1993). For a mathematical definition of R_0 for homogeneous operators we refer to Thieme et al. (2016) and Jin and Thieme (2016).

From a conceptual biological point of view, it makes sense to think of R_0 as a function of two variables, viz. the type/trait of the individuals and the condition of the environment. The linear setting assumes that individuals have independent lives, but this clearly is an idealization if population growth leads to increasing numbers. Density dependence arises by feedback to the environmental condition (as a concrete example, think of consumption of food). Often this results in a steady state : the environmental condition is set such that R_0 equals one. One can now introduce in low quantity another type of individual and ask whether its population will grow. Thus consideration of R_0 enters in the analysis of competition models. Often this leads to statements about optimization of R_0, but the mechanistically more informative formulation is in terms of the pessimization of the environmental condition (Diekmann, 2004; Smith and Thieme, 2013).

As indicated in the main text, computation of R_0 is often facilitated by identifying the possible states at birth, i.e., by studying the range of F. A more detailed analysis of the life cycle graph may help to further reduce the computational burden, see Rueffler and Metz (2013), de Camino-Beck et al. (2009) and the references given there.

References

Bacaër, N., Ait Dads, E.H., 2012. On the biological interpretation of a definition for the parameter R_0 in periodic population models. J. Math. Biol. 65, 601–621.

Bani-Yaghoub, M., Gautam, R., Shuai, Z., van den Driessche, P., Ivanek, R., 2012. Reproduction numbers for infections with free-living pathogens growing in the environment. J. Biol. Dyn. 6, 923–940.
Berman, A., Plemmons, R.J., 1994. Nonnegative Matrices in the Mathematical Sciences, Classics in Applied Mathematics No. 9, SIAM, Philadelphia,.
Browne, C., Webb, G.F., 2015. A nosocomial epidemic model with infection of patients due to contaminated rooms. Math. Biosci. Eng. 12 (4), 761–787.
Caswell, H., 2001. Matrix Population Models: Construction, Analysis and Interpretation, second ed. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.
Cushing, J.M., Ackleh, A.S., 2012. A net reproductive number for periodic matrix models. J. Biol. Dyn. 6, 166–188.
Cushing, J.M., Yicang, Z., 1994. The net reproductive value and stability in matrix population models. Nat. Resour. Model. 8, 297–333.
Cushing, J.M., 1998. An Introduction to Structured Population Dynamics, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 71, SIAM, Philadelphia.
de Camino-Beck, T., Lewis, M.A., van den Driessche, P., 2009. Graph-theoretic method for the basic reproduction number in continuous time epidemiological models. J. Math. Biol. 59 (4), 503–516.
Diekmann, O., Heesterbeek, J.A.P., Roberts, M.G., 2010. The construction of next-generation matrices for compartmental epidemic models. J. R. Soc. Interface 7, 873–885.
Diekmann, O., Heesterbeek, J.A.P., Britton, T., 2013. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, Princeton, USA, and Oxford, UK.
Diekmann, O., 2004. A Beginner's Guide to Adaptive Dynamics. 63. Banach Center Publication, pp. 47–86.
Eager, E.A., Rebarber, R., Tenhumberg, B., 2014. Global asymptotic stability of plant-seed bank models. J. Math. Biol. 69, 1–13.
Heesterbeek, J.A.P., Metz, J.A.J., 1993. The saturating contact rate in marriage- and epidemic models. J. Math. Biol. 31 (5), 529–539.
Heesterbeek, J.A.P., Roberts, M.G., 2007. The type-reproduction number T in models for infectious disease control. Math. Biosci. 206, 3–10.
Heesterbeek, J.A.P., 2002. A brief history of R_0 and a recipe for its calculation. Acta Biotheor. 50, 189–204. http://dx.doi.org/10.1023/A:1016599411804.
Heffernan, J.M., Smith, R.J., Wahl, L.M., 2005. Perspectives on the basic reproductive ratio. J. R. Soc. Interface 2, 281–293. http://dx.doi.org/10.1098/rsif.2005.0042.
Inaba, H., Nishiura, H., 2008. The state-reproduction number for a multistate class age structured epidemic system and its application to the asymptomatic transmission model. Math. Biosci. 216, 77–89.
Jin, W., Thieme, H.R., 2016. An extinction/persistence threshold for sexually reproducing populations: the cone spectral radius. Discrete Contin. Dyn. Syst. B 21, 447–470. http://dx.doi.org/10.3934/dcdsb.2016.21.447.
Jin, W., Smith, H.L., Thieme, H.R., 2015. Persistence versus extinction for a class of discrete-time structured population models. J. Math. Biol. 02 (June). http://dx.doi.org/10.1007/s00285-015-0898-8, online.
Keeling, M.J., Grenfell, B.T., 2000. Individual-based perspectives on R_0. J. Theor. Biol. 203, 51–61. http://dx.doi.org/10.1006/jtbi.1999.106.
Li, C.-K., Schneider, H., 2002. Applications of Perron–Frobenius theory to population dynamics. J. Math. Biol. 44, 450–462.
Li, J., Blakeley, D., Smith?, R.J., 2011. The failure of R_0. Computational and Mathematical Methods in Medicine. Article ID 527610. http://dx.doi.org/10.1155/2011/527610.
Roberts, M.G., Heesterbeek, J.A.P., 2003. A new method for estimating the effort required to control an infectious disease, in: Proceedings of the Royal Society, London B, vol. 270, pp. 1359–1364.
Roberts, M.G., 2007. The pluses and minuses of R_0. J. R. Soc. Interface 4, 949–961, URL ⟨http://rsif.royalsocietypublishing.org/content/4/16/949⟩.
Rueffler, C., Metz, J.A.J., 2013. Necessary and sufficient conditions for R_0 to be a sum of contributions of fertility loops. J. Math. Biol. 66, 1099–1122.
Smith, H.L., Thieme, H.R., 2013. Chemostats and epidemics: competition for nutrients/hosts. Math. Biosci. Eng. 10, 1635–1650.
Thieme, H.R., 2016. Spectral radii and Collatz-Wielandt numbers for homogeneous order-preserving maps and the monotone companion norm. In: de Jeu, M., de Pagter, B., van Gaans, O., Veraar, M. (Eds.) Ordered Structures and Applications. Birkhäuser, Springer International Publishing, Switzerland (in press), http://dx.doi.org/10.1007/978-3-319-27842-1.
Thieme, H.R., 2009. Spectral bound and reproduction number for infinite dimensional population structure and time-heterogeneity. SIAM J. Appl. Math. 70, 188–211.
Wang, W., Zhao, X.-Q., 2008. Thereshold dynamics for compartmental epidemic models in periodic environments. J. Dyn. Differ. Equ. 20, 699–717.