

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/230684314>

On the formulation of epidemic models (an appraisal of Kermack and McKendrick)

Article in *Journal of Biological Dynamics* · August 2012

DOI: 10.1080/17513758.2012.716454 · Source: PubMed

CITATIONS

36

READS

373

5 authors, including:



Dimitri Breda

University of Udine

61 PUBLICATIONS 931 CITATIONS

[SEE PROFILE](#)



Andrea Pugliese

Università degli Studi di Trento

118 PUBLICATIONS 1,649 CITATIONS

[SEE PROFILE](#)



Rossana Vermiglio

University of Udine

70 PUBLICATIONS 1,254 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



LEXEM - Laboratory of excellence for epidemiology and modeling [View project](#)

On the formulation of epidemic models (an appraisal of Kermack and McKendrick)

D. Breda¹, O. Diekmann², W.F. de Graaf², A. Pugliese³, R. Vermiglio¹

¹Dipartimento di Matematica e Informatica, Università degli Studi di Udine, Udine, Italy

²Department of Mathematics, Utrecht University, Utrecht, The Netherlands

³Dipartimento di Matematica, Università di Trento, Povo (TN), Italy

Dedicated to Simon Levin at the occasion of his 70th birthday

Thank you Simon for being an organizing center for mathematical biology and a source of inspiration for two of the authors for many years

1 Introduction

As early as 1927 W.O. Kermack and A.G. McKendrick published a paper “A Contribution to the Mathematical Theory of Epidemics” in the Proceedings of the Royal Society London Ser. A [1]. The paper became a classic in infectious disease epidemiology and has been cited innumerable times. It was reprinted, with a discussion by Roy Anderson, in a special issue “Classics of Theoretical Biology” (part two) in the Bulletin of Mathematical Biology [2].

But how often is it actually read? Judging from an incessant misconception of its contents, one is inclined to conclude: hardly ever! Indeed, even experienced experts often believe that the paper is just about the system

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \alpha I \\ \frac{dR}{dt} &= \alpha I\end{aligned}$$

of three ODEs. This system figures prominently in Section 3.2 of the paper, but there it is explicitly described as a very special (and important) case of a much more general model formulated in Section 2. The key feature of the general model is that it incorporates a “per capita rate of infectivity ϕ ” which depends on the time θ elapsed since infection took place. The mathematical incarnation of the general model is a nonlinear renewal equation! A more detailed presentation of the underlying individual-based assumptions and of the derivation of the renewal equation can be found in [3].

The same issue of the Bulletin of Mathematical Biology also reprints two follow-up papers:

- Contributions to the mathematical theory of epidemics-II. The problem of endemicity [4] [5].
- Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity [6] [7].

These follow-up papers are not nearly as much cited as the 1927 paper [1] [2]. A possible explanation might be that the 1927 paper arrives at a general and robust conclusion (viz., the Threshold Theorem) while, in contrast, the other two papers discuss a multitude of formulations and partial results that do not culminate in one robust ‘law’. By the way, see [8] for a recent critique of one particular aspect of the 1927 paper.

In the spirit of [9], the aim of the present paper is to take up “the problem of endemicity” and to

show that, actually, one can formulate it in a way that is very reminiscent of the general epidemic model of 1927, provided one takes the force of infection as the primary unknown. In particular we shall show that one can derive a **scalar** nonlinear renewal equation for the force of infection under rather general assumptions on both the demographic turnover and the waning of immunity. In order to introduce this approach we first rederive in Section 2 the main results of the 1927 paper by employing the formulation in terms of the renewal equation for the force of infection. In this “epidemic” setting the population is demographically closed and infection leads to permanent immunity. In Section 3 we relax the first of these assumptions and introduce newborn individuals, which are susceptible, at a constant rate B . Rather than introducing a per capita death rate, we describe the survival probability till at least age a by $\mathcal{F}(a)$ in order to be able to incorporate the rather flat age distributions that characterize modern developed countries. We derive the unsurprising result that there is a unique endemic equilibrium when $R_0 > 1$ and no endemic equilibrium when $R_0 < 1$. For the special case of a constant per capita death rate μ (so $\mathcal{F}(a) = e^{-\mu a}$) we show that the endemic equilibrium is locally asymptotically stable, no matter how infectiousness depends on time since infection. We are unable to extend this result to general survival functions and wonder whether or not one might have the possibility of a Hopf bifurcation in that setting?

In Section 4 we drop the assumption of permanent immunity and allow for re-infection of the same individual. We note that there are multiple options for describing partial protection but that, whichever of these we choose, the force of infection is determined by a scalar renewal equation (the difference is in the kernel). We briefly discuss the existence and uniqueness of an endemic state, but essentially leave this as an open problem (for ourselves) and we do not even start to discuss the question of local stability. Our plan is to investigate these issues in the near future. Moreover, we want to incorporate vaccination schedules and multiple strains (perhaps in Adaptive Dynamics spirit) in order to study how (and on what time scale) vaccination may lead to strain replacement in populations with realistic demographic structure.

In an appendix we show how the SIS compartmental model fits into the framework described in this paper.

2 An epidemic in a closed population

We denote

$$\begin{aligned} F(t) &:= \text{force of infection at time } t \\ S(t) &:= \text{density (number per unit area) of susceptibles at time } t \end{aligned}$$

and recall that the force of infection is, by definition, the probability per unit of time that a susceptible becomes infected. So if numbers are large enough to warrant a deterministic description, we have

$$\text{incidence} = F(t)S(t)$$

where “incidence” is defined as the number of new cases per unit of time and area. In a demographically closed population the variable S only changes due to transmission of infection, i.e.,

$$\dot{S}(t) = -\text{incidence}$$

if, as indeed we assume, the infection leads to permanent immunity. As the key modelling ingredient we now introduce

$$\begin{aligned} A(\tau) &:= \text{expected contribution to the force of infection by an individual that was itself} \\ &\quad \text{infected } \tau \text{ units of time ago} \end{aligned}$$

The constitutive equation

$$F(t) = \int_0^\infty F(t-\tau)S(t-\tau)A(\tau)d\tau \tag{2.1}$$

now tells us how the current force of infection depends on past incidence. By integrating

$$\dot{S}(t) = -F(t)S(t)$$

we obtain

$$S(t) = S(-\infty)e^{-\int_{-\infty}^t F(\sigma)d\sigma} \quad (2.2)$$

and if we substitute (2.2) into (2.1) we obtain a nonlinear scalar renewal equation for the unknown F .

Quite in general, a delay equation is a rule for extending a function of time towards the future on the basis of the (assumed to be) known past. Such equations lead to dynamical systems by considering the shift along the extended function, see [10] [11]. The renewal equation (2.1), with (2.2) substituted, fits into this framework!

But before we rederive the main conclusion of the Kermack-McKendrick 1927 paper from (2.1), let us briefly discuss how one would choose the function A of time-since-infection τ .

First of all we emphasize that familiar compartmental epidemic models are included in the present framework:

$$\begin{aligned} A(\tau) &= \beta e^{-\alpha\tau} \Leftrightarrow \text{SIR} \\ A(\tau) &= \beta \frac{\gamma}{\gamma - \alpha} (e^{-\alpha\tau} - e^{-\gamma\tau}) \Leftrightarrow \text{SEIR} \end{aligned}$$

(we leave it to our readers to elaborate the \Leftrightarrow ; the general idea that delay equations with kernels defined

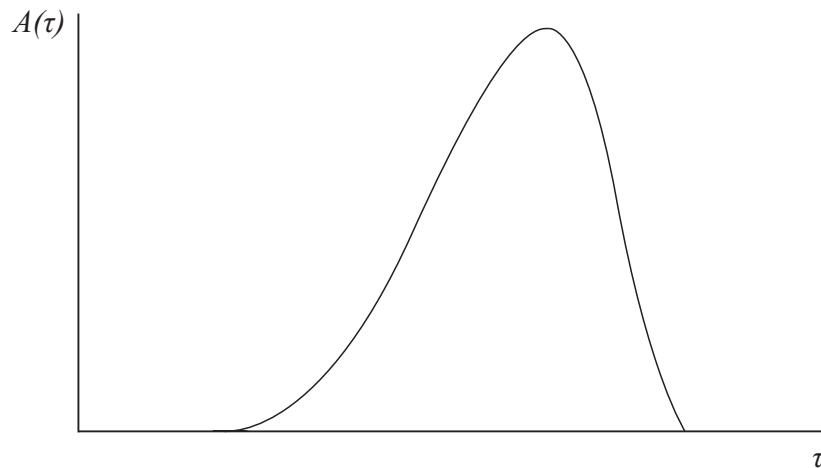


Figure 1: A typical infectivity function $A(\tau)$

in terms of matrix exponentials correspond to systems of ODE is called “Linear Chain Trickery”, see e.g. [12]). So certain parametrized families of A correspond to well-known ODE models and the parameters can thus be interpreted like the mean duration of the infectious period etcetera. Quite generally, A has two components:

- contact intensity
- infectiousness (i.e., probability of transmission, given a contact with a susceptible)

Detailed medical information may yield an idea of relative infectiousness as a function of τ . The specification of one scalar factor is then needed to turn this into a graphical specification of A , see Figure 1. As we will see below, some indicators of “severity” (in particular R_0 and the final size) depend

only on the integral of A and not on the form of the graph of A .

In order to be mathematically precise we assume:

$$\begin{aligned} A(\tau) &\geq 0 \\ A : [0, \infty) &\rightarrow [0, \infty) \text{ is integrable} \end{aligned}$$

Our aim is now to derive a general conclusion from (2.1) and (2.2). To do so, we introduce the cumulative force of infection

$$y(t) := \int_{-\infty}^t F(\sigma) d\sigma \quad (2.3)$$

By integrating (2.1) from $-\infty$ to t , interchanging the order of the integrals and using

$$F(t - \tau)S(t - \tau) = -\dot{S}(t - \tau)$$

and then (2.2), we obtain the scalar nonlinear renewal equation

$$y(t) = \int_0^\infty (1 - e^{-y(t-\tau)}) S(-\infty) A(\tau) d\tau \quad (2.4)$$

(note, incidentally, that (2.4) is of convolution type whereas (2.1), with (2.2) inserted, is not). Since $F \geq 0$ we know that y should be an increasing function of t and with a little bit of effort one can check that the rule for extension (2.4) does indeed preserve this property, i.e. if y is increasing on $(-\infty, 0)$ then y is increasing on $(0, +\infty)$ [13]. Moreover, since $1 - e^{-y} \leq 1$ for $y \geq 0$ and A is integrable, y is bounded. So the limit $y(\infty) = \lim_{t \rightarrow \infty} y(t)$ exists. Using again the integrability of A one shows that $y(\infty)$ has to satisfy the equation

$$y(\infty) = R_0(1 - e^{-y(\infty)}) \quad (2.5)$$

where

$$R_0 := S(-\infty) \int_0^\infty A(\tau) d\tau \quad (2.6)$$

can be interpreted as the expected number of secondary cases caused by a primary case introduced in a population with susceptible density $S(-\infty)$ or, in the standard jargon, R_0 is the *basic reproduction ratio*.

Using elementary properties of the function $y \mapsto 1 - e^{-y}$ and simple graphical arguments, one proves that (2.5) has a unique strictly positive solution if $R_0 > 1$ and no strictly positive solution if $R_0 \leq 1$. In order to translate this information into the Kermack-McKendrick threshold theorem, we first observe that (2.2) implies

$$1 - \frac{S(\infty)}{S(-\infty)} = 1 - e^{-y(\infty)} = \frac{1}{R_0} y(\infty) \quad (2.7)$$

where the left hand side represents the *final size* of the epidemic, i.e., the fraction of the population that is infected, sooner or later, during an outbreak. Thus we are led to conclude:

- i. when $R_0 > 1$, introduction of the infective agent leads to an outbreak with final size given by (2.7), where $y(\infty)$ is the strictly positive solution of (2.5)
- ii. when $R_0 \leq 1$, introduction of the infective agent leads to an outbreak with final size close to zero

But several comments are essential for both the derivation of this conclusion and its interpretation:

Concerning i.: this is the deterministic formulation that takes as a starting point that a small but positive *fraction* of the large population is infected. When we start with a small *number* of infected

individuals, we should describe the initial stages by a branching process. The condition $R_0 > 1$ then amounts to the branching process being supercritical. As a branching process may go extinct, even when supercritical, we may get a *minor outbreak*. Equation (2.7) describes the expected size of a major outbreak. See [14].

Concerning ii.: For $R_0 \leq 1$ the relevant solution of (2.7) is zero, so where does the “close to zero” come from? In deriving (2.5) we assumed that y satisfied (2.4) for all $t \in (-\infty, \infty)$. When providing an initial condition that captures the introduction of a small number of infectives, we essentially replace (2.4) by

$$y(t) = \int_0^t (1 - e^{-y(t-\tau)}) S(-\infty) A(\tau) d\tau + f(t) \quad (2.8)$$

where $f(t)$ is a given function. Thus we obtain

$$y(\infty) = R_0(1 - e^{-y(\infty)}) + f(\infty) \quad (2.9)$$

and conclusion ii. should be read as follows:

ii'. when $R_0 \leq 1$, the positive solution $y(\infty)$ of (2.9) converges to zero when $f(\infty) \downarrow 0$

See Exercise 1.12.iv in [14] for details.

If transmission happens on a time scale which is fast relative to the time scale of demographic changes (birth, death, migration), we may conceive the host population as static during the outbreak. If, moreover, infection confers permanent immunity, the fraction of the population that is susceptible necessarily decreases monotonically and the concept of “final size” makes sense. In this section we have shown that the expected final size can be characterized in terms of the scalar equation (2.5) involving only one parameter, the basic reproduction number R_0 . In turn, R_0 is defined by (2.6), i.e., as the product of the original density of susceptibles and the expected total contribution to the force of infection of one newly infected individual. The equation (2.5) was derived from a nonlinear scalar renewal equation.

The entire content of this section is contained in the 1927 paper of Kermack-McKendrick! The only thing we have done is to rewrite the derivation in such a way that the transition to the next sections, dealing with demographic turnover and with waning immunity, is as smooth as possible.

3 The endemic balance

Even if individuals remain immune for the rest of their life after an infection, new susceptibles are produced by reproduction. In this section we consider the situation that, at the population level, there is a constant birth rate B . So from the point of view of the infectious agent, there is a constant inflow at rate B of “resource”.

Individuals come and go: we now also need to specify how long newborn individuals live. We do so in terms of the probability $\mathcal{F}(a)$ that an individual stays alive for at least a units of time. In particular we assume that this survival probability does not depend on whether or not the individual becomes infected (so we are investigating the influence of demographic turnover on disease transmission and not the influence of a deadly infectious disease like HIV on population dynamics). We want to keep \mathcal{F} general and not restrict to the special family $\mathcal{F}(a) = e^{-\mu a}$ with $\mu > 0$, simply since in modern developed countries the observed \mathcal{F} is hugely different from exponential. Note that we have introduced two more model ingredients (on top of $A(\tau)$), viz., B and $\mathcal{F}(a)$. We assume that $\mathcal{F}(0) = 1$, that \mathcal{F} is monotonically non-increasing and that $\mathcal{F} \rightarrow 0$ exponentially for $a \rightarrow \infty$. We consider $B > 0$.

If no infectious disease is circulating, we observe the stable age distribution

$$S(t, a) = B\mathcal{F}(a)$$

where S is now the number of susceptibles per unit of area and age. If we assume that “age” has no impact on susceptibility, nor on the force of infection that an individual experiences, we obtain

$$S(t, a) = B\mathcal{F}(a)e^{-\int_0^a F(t-a+\sigma)d\sigma} \quad (3.1)$$

when $F(t)$ is the prevailing force of infection at time t (indeed, an individual that has age a at time t experienced a force of infection $F(t-a+\sigma)$ at age σ). If we substitute (3.1) into

$$\text{incidence} = F(t) \int_0^\infty S(t, a)da$$

we have, exactly as before, expressed the incidence in terms of past values of the force of infection. Instead of the constitutive equation (2.1) we now postulate

$$F(t) = \int_0^\infty F(t-\tau) \int_0^\infty S(t-\tau, a) \frac{\mathcal{F}(a+\tau)}{\mathcal{F}(a)} da A(\tau) d\tau \quad (3.2)$$

since we interpret A as an expected contribution given survival, and so the true expected contribution has the conditional survival probability $\mathcal{F}(a+\tau)/\mathcal{F}(a)$ as an extra factor.

If we substitute (3.1) into (3.2) we arrive at a nonlinear scalar renewal equation for the unknown F . Apart from the disease free state $F(t) \equiv 0$ there might be an endemic steady state. To find it, we need to solve

$$1 = B \int_0^\infty \int_0^\infty e^{-F a} \mathcal{F}(a+\tau) da A(\tau) d\tau \quad (3.3)$$

The right hand side is a monotonically decreasing function of F , converging to zero for $F \rightarrow \infty$. So if we assume that the value at $F = 0$, which we denote by R_0 (since it is indeed the expected number of secondary cases caused by a primary case introduced in a susceptible population with age distribution $B\mathcal{F}(a)$), exceeds one, there exists exactly one steady endemic force of infection. And if $R_0 < 1$, no such steady endemic force of infection does exist.

To ascertain the local stability of the endemic steady state, we linearize (3.2)-(3.1) and look for solutions of the form $e^{\lambda t}$ for the linearized equation

$$\begin{aligned} u(t) = & B \int_0^\infty u(t-\tau) \int_0^\infty e^{-F a} \mathcal{F}(a+\tau) da A(\tau) d\tau \\ & - BF \int_0^\infty \int_0^\infty e^{-F a} \mathcal{F}(a+\tau) \int_0^a u(t-\tau-a+\sigma) d\sigma da A(\tau) d\tau. \end{aligned}$$

This leads to the *characteristic equation*

$$1 = B \int_0^\infty \int_0^\infty e^{-F a} \mathcal{F}(a+\tau) \left(1 - \frac{F}{\lambda} (1 - e^{-a\lambda})\right) e^{-\lambda\tau} da A(\tau) d\tau \quad (3.4)$$

with F the unique positive solution of (3.3) for $R_0 > 1$. It is an open problem to determine whether or not (3.4) has solutions with $\text{Re}(\lambda) > 0$ for general \mathcal{F} . Remarkably, we can show that when $\mathcal{F}(a) = e^{-\mu a}$ all roots satisfy $\text{Re}(\lambda) < 0$ (for **any** non-negative and integrable function A !).

Notation: \bar{A} is the Laplace transform of A , i.e., $\bar{A}(z) := \int_0^\infty A(\tau)e^{-z\tau} d\tau$.

Theorem 3.1. Assume that $\mathcal{F}(a) = e^{-\mu a}$ with $\mu > 0$ and that $R_0 > 1$ where

$$R_0 = \frac{B}{\mu} \bar{A}(\mu)$$

The solution

$$F = B\bar{A}(\mu) - \mu \quad (3.5)$$

of (3.3) is locally asymptotically stable as a steady state of (3.2)-(3.1).

Proof. By the theory developed in [11] [18], it suffices to show that all roots of (3.4) have negative real part. A straightforward computation, using also (3.5), shows that (3.4) takes the form

$$1 = \frac{\lambda + \mu}{\lambda + B\bar{A}(\mu)} \frac{\bar{A}(\lambda + \mu)}{\bar{A}(\mu)} \quad (3.6)$$

when $\mathcal{F}(a) = e^{-\mu a}$. Since (recall that $A(\tau) \geq 0$)

$$|\bar{A}(\lambda + \mu)| \leq \bar{A}(\operatorname{Re}(\lambda) + \mu)$$

and, for $\operatorname{Re}(\lambda) \geq 0$

$$\bar{A}(\operatorname{Re}(\lambda) + \mu) \leq \bar{A}(\mu)$$

the absolute value of the second factor at the right hand side is bounded above by one. Since $R_0 > 1$ we have

$$B\bar{A}(\mu) + \operatorname{Re}(\lambda) > \mu + \operatorname{Re}(\lambda) > 0$$

for $\operatorname{Re}(\lambda) \geq 0$ and hence

$$\left| \frac{\lambda + \mu}{\lambda + B\bar{A}(\mu)} \right| = \frac{\sqrt{(\mu + \operatorname{Re}(\lambda))^2 + (\operatorname{Im}(\lambda))^2}}{\sqrt{(B\bar{A}(\mu) + \operatorname{Re}(\lambda))^2 + (\operatorname{Im}(\lambda))^2}} < 1$$

So for $\operatorname{Re}(\lambda) \geq 0$ the right hand side of (3.6) is in absolute value less than one and consequently (3.6) cannot hold for such λ . \square

Remark. The proof above is a much simplified version of the elaboration of Exercise 3.10 in [14].

In this section we have shown that when demographic turnover is described by a constant population birth rate and a general survival function, we can still capture the infectious disease dynamics by a scalar renewal equation for the force of infection. This equation has a positive steady state if and only if $R_0 > 1$ and there is at most one such steady state. For the special case of a constant per capita death rate we find that this steady state is locally asymptotically stable. Whether or not it is for general survival is an open problem. But please note that [15] [16] [17] find that instability is certainly possible if we allow the susceptibility of an individual to depend on its age.

4 Waning immunity

When immunity reduces susceptibility, but has no impact on infectiousness once infection took place, we can stick to the already introduced $A(\tau)$ to describe the expected contribution to the force of infection at time τ after infection, given survival.

Concerning temporary partial immunity, there are at least two ways to describe it. One may assume that all individuals, at time τ after the last infection took place, have their susceptibility reduced by a factor $Q(\tau)$, meaning that the probability of getting infected upon contact with an infectious individual is, $Q(\tau)$ times what it is for a fully susceptible individual (so $Q(\tau) = 0$ corresponds to full protection and $Q(\tau) = 1$ to full susceptibility; incidentally, Inaba [9] follows Kermack and McKendrick [4] [5] [6] [7] in assuming that a new clock starts when an event transforms an infected individual into a recovered individual and that the reduction in susceptibility is a function of the time shown by this second clock). Another possibility is to allow only for full protection and full susceptibility and to describe the probability that at time τ after the last infection an individual is fully susceptible again by $P(\tau)$. Note that SIS and SIRS compartment models (see Appendix) are of this type. Also note that

one can of course introduce a compartment of intermediate protection (see e.g. A. Swart e.a. [19]) and thus obtain a description that combines features of both gradual susceptibility return and variability in the length of the time window of protection. One might also assume that susceptibility does not simply depend on the time since last infection, but also on the history of previous infections.

Note that we have to make sure that an infected individual that contributes to positivity of $P(\tau)$ should **not** contribute to positivity of $A(\sigma)$ for $\sigma \geq \tau$, as otherwise the possibility of re-infection does have an influence on expected infectiousness (for compartmental models of SIS and SIRS type etcetera this is indeed guaranteed). When we work with Q , the safest assumption is to require that $Q(\tau) = 0$ for $\tau \in \text{supp}(A)$.

In any case we can assume that equation (3.2), which we here repeat as

$$F(t) = \int_0^\infty F(t-\tau) \int_0^\infty S(t-\tau, a) \frac{\mathcal{F}(a+\tau)}{\mathcal{F}(a)} da A(\tau) d\tau \quad (4.1)$$

remains valid, interpreting (in case of reduced susceptibility) $S(t, a)$ in terms of susceptible equivalents, in the sense that one individual with reduced susceptibility $Q(\tau)$ is counted as $Q(\tau)$ susceptibles. Now we have to change the expression (3.1) for the number of susceptibles per unit of area and age, in order to account for regained susceptibility (long) after infection.

Another important observation is that, implicitly, P and Q refer to an individual while conditioning that this individual is not yet re-infected. The possibility of re-infection leads, in the case of partial susceptibility, to the conclusion that of the individuals that become infected at time $t - \tau$ a fraction

$$e^{-\int_0^\tau Q(\sigma) F(t-\tau+\sigma) d\sigma}$$

will not have been reinfected before time t . If we work in the P -framework and consider again an individual infected at time $t - \tau$, the probability that it is susceptible at time t is given by the Stieltjes integral

$$\int_0^\tau e^{-\int_0^{\tau-\sigma} F(t-\tau+\sigma+\eta) d\eta} P(d\sigma)$$

(indeed, if the individual regains full susceptibility at time σ after infection, it has to escape from re-infection on a time interval of length $\tau - \sigma$ beginning at $t - \tau + \sigma$). These observations motivate us to define

$$\mathcal{G}(\tau, F_t) := \begin{cases} \text{either} & Q(\tau) e^{-\int_0^\tau Q(\sigma) F(t-\tau+\sigma) d\sigma} \\ \text{or} & \int_0^\tau e^{-\int_0^{\tau-\sigma} F(t-\tau+\sigma+\eta) d\eta} P(d\sigma) \end{cases} \quad (4.2)$$

in order to capture Q -models and P -models in one uniform notation. Here we use the notational convention

$$F_t(\sigma) := F(t + \sigma), \sigma \leq 0 \quad (4.3)$$

of the theory of delay equations [10] [20].

We are now ready to replace the expression (3.1) by the equation

$$S(t, a) = B\mathcal{F}(a) e^{-\int_0^a F(t-a+\sigma) d\sigma} + \int_0^a F(t-\tau) S(t-\tau, a-\tau) \frac{\mathcal{F}(a)}{\mathcal{F}(a-\tau)} \mathcal{G}(\tau, F_t) d\tau \quad (4.4)$$

The model is now described by the two equations (4.1) and (4.4) for the two unknowns $F(t)$ and $S(t, a)$ and it has as its ingredients B , $\mathcal{F}(a)$, $A(\tau)$ and either $Q(\tau)$ or $P(\tau)$. For given F the equation (4.4) is linear in S , so solving (4.4) is in principle straightforward and upon substitution of the result into (4.1) we end up with a scalar nonlinear renewal equation for F . In fact one can solve (4.4) by successive approximation, which amounts to generation expansion: the term $B\mathcal{F}(a) e^{-\int_0^a F(t-a+\sigma) d\sigma}$ corresponds to individuals that have never been infected, the next term

$$B\mathcal{F}(a) \int_0^a F(t-\tau) e^{-\int_0^{a-\tau} F(t-a+\sigma) d\sigma} \mathcal{G}(\tau, F_t) d\tau$$

to individuals that have been infected once, etcetera. Whenever the support of Q or P is bounded away from zero (meaning that the time window between two successive infections has a strictly positive minimal length), there are, at any given finite age, only finitely many non-zero terms in the generation expansion. So if in addition $\mathcal{F}(a) = 0$ for any $a \geq a_{\max}$, we find in this way an exact representation of the solution of (4.4) as a finite sum of explicit expressions involving the history of F . We conclude that the dynamics of an epidemic model that incorporates a general survival function $\mathcal{F}(a)$ and various quite general forms of waning immunity, is still fully described by a **scalar** nonlinear renewal equation!

If we define $q(t, a)$ by the relation

$$S(t, a) = B\mathcal{F}(a)q(t, a) \quad (4.5)$$

then q describes the relative susceptibility (Q -version) or the probability to be susceptible (P -version) of an individual of age a at time t , given survival. Upon substitution of this relation into (4.4) we find for q the equation

$$q(t, a) = e^{-\int_0^a F(t-a+\sigma)d\sigma} + \int_0^a F(t-\tau)q(t-\tau, a-\tau)\mathcal{G}(\tau, F_t)d\tau \quad (4.6)$$

while substitution into (4.1) leads to

$$F(t) = B \int_0^\infty F(t-\tau) \int_0^\infty q(t-\tau, a)\mathcal{F}(a+\tau)daA(\tau)d\tau \quad (4.7)$$

Clearly $F(t) \equiv 0$, $q(t, a) \equiv 1$ corresponds to the disease free steady state. An endemic steady state is characterized by

$$q(a) = e^{-Fa} + F \int_0^a q(a-\tau)\mathcal{G}(\tau, F)d\tau \quad (4.8)$$

$$1 = B \int_0^\infty \int_0^\infty q(a)\mathcal{F}(a+\tau)daA(\tau)d\tau \quad (4.9)$$

The right hand side of (4.9) for $q(a) \equiv 1$ is again equal to R_0 , just as it was in the preceding section (the possibility of re-infection has no impact on R_0 for the simple reason that when infection is very rare, it can be excluded that the same individual is infected again). If, as before, we solve the linear equation (4.8) by generation expansion and substitute the result into (4.9), we obtain the scalar equation from which we have to determine the unknown $F > 0$.

In the special case that $\mathcal{F}(a) = e^{-\mu a}$ equation (4.9) reduces to

$$1 = B\bar{A}(\mu)\bar{q}(\mu) \quad (4.10)$$

Equation (4.8) is a linear renewal equation, so it can be solved by Laplace transformation, leading to

$$\bar{q}(\mu) = \frac{1}{1 - F\bar{\mathcal{G}}(\mu, F)} \frac{1}{\mu + F} \quad (4.11)$$

In the P -version we have

$$\bar{\mathcal{G}}(\mu, F) = \frac{1}{\mu + F} \bar{P}(\mu) \quad (4.12)$$

where $\bar{P}(\mu)$ is the Laplace-Stieltjes transform of P

$$\bar{P}(\mu) := \int_0^\infty e^{-\mu\tau} P(d\tau) \quad (4.13)$$

So for the P -version we find

$$\bar{q}(\mu) = \frac{1}{\mu + F(1 - \bar{P}(\mu))} \quad (4.14)$$

and since $\bar{P}(\mu) < 1$ the right hand side is a decreasing function of F . Accordingly (4.10) has a unique positive solution whenever $R_0 > 1$ and no such solution when $R_0 \leq 1$. We conjecture that the same is true for the Q -version, but so far we did not manage to prove this. We also conjecture that this conclusion holds for general $\mathcal{F}(a)$. In order to show this, we might rewrite the factor

$$\int_0^\infty q(a)\mathcal{F}(a+\tau)da$$

in (4.9) by integration by parts as

$$-\int_0^\infty \int_0^a q(\alpha)d\alpha\mathcal{F}'(a+\tau)da$$

(where, for convenience, we have assumed that \mathcal{F} is differentiable; since \mathcal{F} is anyhow monotone decreasing, the general case can be dealt with by using a Stieltjes integral). Now note that $\int_0^a q(\alpha)d\alpha$ is the expected time that an individual that survived till age a spent susceptible (weighed in the appropriate manner when we work with the Q -version). So it seems very reasonable to conjecture that

$$\frac{d}{dF} \int_0^a q(\alpha)d\alpha < 0$$

(note that q itself may depend in a more complicated way on F , as an increase in F leads first to a decrease of the expected age of first infection but next to an increase of *return* to the susceptible class at an early age). If we define

$$\chi(a) := \int_0^a q(\alpha)d\alpha \tag{4.15}$$

we obtain by integration of (4.8) the linear renewal equation

$$\chi(a) = \frac{1 - e^{-Fa}}{F} + F \int_0^a \chi(a-\tau)\mathcal{G}(\tau, F)d\tau \tag{4.16}$$

It remains to study in detail how the solution of this renewal equation depends on the parameter F .

5 Concluding Remarks

Often when epidemic models are formulated as delay equations, the aim is to derive conditions for the (in)stability of the endemic steady state and to find out, by way of the Hopf bifurcation theorem, when one should expect to find persistent oscillations [17] [9] [21] [22] [23] [24] [25] [26] [27].

The trigger for the approach sketched in the present paper is somewhat different: our ultimate goal is to find conditions for strain replacement as a result of mass vaccination. In this context the need arises to incorporate

- i. a realistic form of demographic turnover
- ii. the waning of immunity
- iii. cross-immunity

The text above describes our approach for dealing with issues i. and ii. We realise that dealing in an effective yet meaningful way with issue iii. presents a major challenge. Advice of Simon Levin on this matter will be highly appreciated!

A Appendix

In this appendix we show how the familiar SIS compartmental model described by the ODE system

$$\begin{aligned}\frac{dS}{dt} &= B - \mu S - \beta IS + \alpha I \\ \frac{dI}{dt} &= -\mu I + \beta IS - \alpha I\end{aligned}\tag{A.1}$$

fits into the framework described in this paper. Hopefully our presentation is such that the reader is convinced that also models with additional exposed and/or removed compartments, such as the SEIRS model described by

$$\begin{aligned}\frac{dS}{dt} &= B - \mu S - \beta IS && + \delta R \\ \frac{dI}{dt} &= -\mu E + \beta IS - \gamma E \\ \frac{dI}{dt} &= -\mu I && + \gamma E - \alpha I \\ \frac{dR}{dt} &= -\mu R && + \alpha I - \delta R\end{aligned}\tag{A.2}$$

are covered by our approach.

Let for a given function $F = F(t)$ the two-vector $M(t; s)$ be defined by

$$\begin{aligned}\frac{dM_1}{dt}(t; s) &= -F(t)M_1(t; s) + \alpha M_2(t; s), & M_1(s; s) &= 1 \\ \frac{dM_2}{dt}(t; s) &= F(t)M_1(t; s) - \alpha M_2(t; s), & M_2(s; s) &= 0\end{aligned}\tag{A.3}$$

So M is the first column of the fundamental matrix solution of the linear ODE system generated by the matrix

$$\begin{pmatrix} -F(t) & \alpha \\ F(t) & -\alpha \end{pmatrix}$$

and if we define

$$\begin{pmatrix} S \\ I \end{pmatrix}(t) = \int_{-\infty}^t B e^{-\mu(t-s)} M(t; s) ds\tag{A.4}$$

it follows that

$$\begin{aligned}\frac{dS}{dt}(t) &= B - \mu S(t) - F(t)S(t) + \alpha I(t) \\ \frac{dI}{dt}(t) &= -\mu I(t) + F(t)S(t) - \alpha I(t)\end{aligned}\tag{A.5}$$

Our task is to relate equation (4.7) to the equation

$$F(t) = \beta I(t)\tag{A.6}$$

that allows us to identify (A.5) and (A.1). But in order to do so, we need to uncover first the relation between equation (4.6) and the definition of M .

By using for each of the two equations of (A.3) the variation-of-constants formula we obtain

$$\begin{aligned}M_1(t; s) &= e^{-\int_s^t F(\sigma) d\sigma} + \alpha \int_s^t e^{-\int_\sigma^t F(\eta) d\eta} M_2(\sigma; s) d\sigma \\ M_2(t; s) &= \int_s^t e^{-\alpha(t-\sigma)} F(\sigma) M_1(\sigma; s) d\sigma\end{aligned}$$

and upon substitution of the second of these into the first

$$\begin{aligned}
M_1(t; s) &= e^{-\int_s^t F(\sigma) d\sigma} + \alpha \int_s^t e^{-\int_\sigma^t F(\eta) d\eta} \int_s^\sigma e^{-\alpha(\sigma-\theta)} F(\theta) M_1(\theta; s) d\theta d\sigma \\
&= e^{-\int_s^t F(\sigma) d\sigma} + \int_s^t \int_\theta^t e^{-\int_\sigma^t F(\eta) d\eta} \alpha e^{-\alpha(\sigma-\theta)} F(\theta) M_1(\theta; s) d\sigma d\theta \\
&= e^{-\int_s^t F(\sigma) d\sigma} + \int_0^{t-s} F(t-\tau) M_1(t-\tau; s) \int_0^\tau e^{-\int_{t-\tau-\sigma}^t F(\eta) d\eta} \alpha e^{-\alpha\sigma} d\sigma d\tau
\end{aligned}$$

If we now put $M_1(t; s) = q(t, t-s)$ and $s = t-a$ we recover exactly the equation (4.6) for the special case that $P(\tau) = 1 - e^{-\alpha\tau}$ (and hence $P(d\sigma) = \alpha e^{-\alpha\sigma} d\sigma$).

From (A.4) we have

$$I(t) = B \int_0^\infty e^{-\mu a} M_2(t; t-a) da$$

and from the variation-of-constants formula

$$M_2(t; t-a) = \int_{t-a}^t e^{-\alpha(t-\sigma)} F(\sigma) M_1(\sigma; t-a) d\sigma = \int_0^a e^{-\alpha(a-\eta)} F(t-a+\eta) M_1(t-a+\eta; t-a) d\eta$$

If we insert the second of those into the first and write again $M_1(t; s) = q(t, t-s)$, we find upon changing the order of integration

$$\begin{aligned}
I(t) &= B \int_0^\infty \int_\eta^\infty e^{-\mu a - \alpha(a-\eta)} F(t-a+\eta) q(t-a+\eta, \eta) da d\eta \\
&= B \int_0^\infty \int_0^\infty e^{-\mu(\theta+\eta) - \alpha\theta} F(t-\theta) q(t-\theta, \eta) d\theta d\eta \\
&= B \int_0^\infty F(t-\theta) \int_0^\infty q(t-\theta, \eta) e^{-\mu\eta} d\eta e^{-(\mu+\alpha)\theta} d\theta
\end{aligned}$$

If we now multiply both sides by β and insert $A(\tau) = \beta e^{-\alpha\tau}$ and $\mathcal{F}(a) = e^{-\mu a}$ into (4.7), we find that the right hand side of (4.7) coincides with the right hand side of the identity for $\beta I(t)$. In other words, starting from (A.1) we have derived the equation (4.7) for $F(t)$ defined by $F(t) = \beta I(t)$, in the special case that $A(\tau) = \beta e^{-\alpha\tau}$ and $\mathcal{F}(a) = e^{-\mu a}$ and with q defined by (4.6) in the special case that $\mathcal{G}(\tau, F_t) = \int_0^\tau e^{-\int_0^{\tau-\sigma} F(t-\tau+\sigma+\eta) d\eta} \alpha e^{-\alpha\sigma} d\sigma$.

References

- [1] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society London Ser. A* **115**, 700–721 (1927).
- [2] Reprint of W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics in *Classics of Theoretical Biology* (part two), *Bulletin of Mathematical Biology* **53**, 33–55 (1991).
- [3] J.A.J. Metz, The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections, *Acta Biotheoretica* **27**, 75–123 (1978).
- [4] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics-II. The problem of endemicity, *Proceedings of the Royal Society London Ser. A* **138A**, 55–83 (1932).

- [5] Reprint of W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics-II. The problem of endemicity in Classics of Theoretical Biology (part two), *Bulletin of Mathematical Biology* **53**, 57–87 (1991).
- [6] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity, *Proceedings of the Royal Society London Ser. A* **141A**, 94–122 (1933).
- [7] Reprint of W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity in Classics of Theoretical Biology (part two), *Bulletin of Mathematical Biology* **53**, 89–118 (1991).
- [8] N. Bacaër, The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproduction number with seasonality, *J. Math. Biology* **64**, 403–422 (2012).
- [9] H. Inaba, Kermack and McKendrick revisited: The variable susceptibility model for infectious diseases, *Japan J. Indust. Appl. Math.* **18(2)**, 273–292 (2001).
- [10] O. Diekmann, S.A. van Gils, S.M. Verduyn Lunel, H.-O. Walther, Delay Equations: Functional, Complex, and Nonlinear Analysis, volume 110 of Applied Mathematical Sciences, Springer-Verlag (1995).
- [11] O. Diekmann, Ph. Getto, M. Gyllenberg, Stability and bifurcation analysis of Volterra functional equations in the light of suns and stars, *SIAM J. Math. Anal.* **39**, 1023–1069 (2007).
- [12] N. MacDonald, Biological delay systems: linear stability theory, Cambridge University Press, 1989, Ch. 10.
- [13] O. Diekmann, Limiting behaviour in an epidemic model, *NLA-TMA* **1**, 459–470 (1977).
- [14] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation (Mathematical and Computational Biology), John–Wiley (2000).
- [15] H.R. Thieme, Stability change of the endemic equilibrium in age-structured models for the spread of SIR type infectious diseases, *Differential Equations Models in Biology, Epidemiology and Ecology*, S. Busenberg and M. Martelli (eds.), Springer, Berlin, 139–158 (1991).
- [16] V. Andreasen, Instability in an SIR-model with age-dependent susceptibility, *Mathematical Population Dynamics: Analysis of Heterogeneity Volume 1: Theory of Epidemics*, O. Arino, D. Axelrod, M. Kimmel, M. Langlais (eds.), Wuerz Publishing, Winnipeg, 3–14 (1995).
- [17] P. Magal, S. Ruan, Sustained oscillations in an evolutionary epidemiological model of influenza A drift, *Proc. R. Soc. A* **466**, 965–992 (2010).
- [18] O. Diekmann, M. Gyllenberg, Equations with infinite delay: Blending the abstract and the concrete, *Journal of Differential Equations* **252**, 819–851 (2012).
- [19] A.N. Swart, M. Tomasi, M. Kretzschmar, A.H. Havelaar, O. Diekmann, The protective effects of temporary immunity under imposed infection pressure, *Epidemics* **4**, 43–47 (2012).
- [20] J.K. Hale, S.M. Verduyn Lunel, Introduction to Functional Differential Equations, volume 99 of Applied Mathematical Sciences, Springer-Verlag (1993).
- [21] H.W. Hethcote, H.W. Stech, P. Van Den Driessche, Nonlinear Oscillations in Epidemic Models, *SIAM Journal on Applied Mathematics* **40**, 1–9 (1981).
- [22] O. Diekmann, R. Montijn, Prelude to Hopf Bifurcation in an Epidemic Model: Analysis of a Characteristic Equation Associated with a Nonlinear Volterra Integral Equation, *J. Math. Biology* **14**, 117–127 (1982).

- [23] H.W. Hethcote, P. Van Den Driessche, An SIS epidemic model with variable population size and a delay, *J. Math. Biology* **34**, 177–194 (1995).
- [24] K.L. Cooke, P. Van Den Driessche, Analysis of an SEIRS epidemic model with two delays, *J. Math. Biology* **35**, 240–260 (1996).
- [25] K.B. Blyuss, Y.N. Kyrychko, Stability and Bifurcations in an Epidemic Model with Varying Immunity Period, *Bulletin of Mathematical Biology* **72**, 490–505 (2009).
- [26] E. Beretta, D. Breda, An SEIR epidemic model with constant latency time and infectious period, *Mathematical Biosciences and Engineering* **8**, 931–952 (2011).
- [27] G. Huang, E. Beretta, Y. Takeuchi, Global stability for epidemic model with constant latency and infectious periods, *Mathematical Biosciences and Engineering* **9**, 297–312 (2012).