Epidemic fade-out in the Markovian SIR-with-demography infection model

Peter Geoffrey Ballard

March 28, 2018

Thesis by publication,

submitted for the degree of

Doctor of Philosophy

in

Applied Mathematics

at

The University of Adelaide,

Faculty of Engineering, Computer & Mathematical Sciences,

School of Mathematical Sciences.

Contents

Abstract

"Epidemic fade-out" refers to the situation in which an infection is eliminated after an initial major wave of infection. This thesis by publication contains three papers (two published, the third submitted and under review) on the subject of epidemic fade-out in the Markovian SIR-with-demography infection model.

The first paper [6] surveys previous work containing methods for approximating the probability of epidemic fade-out, then proposes a numerical method which is more accurate. Using this method, it surveys trends over a range of parameters, and observes that the probability of epidemic fade-out has a non-monotonic relationship with respect to β, the transmission rate parameter. It shows that this probability often has a local maximum where R_0, the basic reproduction number, is about 2; and gives an explanation for this phenomenon.

The second paper [7] examines the possibility of controlling β, in order to maximise the probability of epidemic fade-out. An optimal policy may be found using Markov decision theory, but this requires very large data structures, meaning this is impractical for all but very small population sizes. So the paper also derives a simple formula for an almost-optimal policy, which can be applied for any population size, and is independent of the values of β.

The third paper [8] extends the Markovian SIR-with-demography infection model to allow β to be time dependent, as the transmission rate may vary with the time of year. It also extends the work to the Markovian SIRS model. It presents an algorithm for calculating the probability of epidemic fade-out for these models, and considers parameters appropriate to influenza-like and measles-like infections. It concludes that the local maximum in the probability of epidemic fade-out is at a value of R_0 somewhat greater than 2, when β is

time-dependent.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.[1]

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed:

Date: 28-3-2018

[1] Both published papers were published with Elsevier [6, 7]. The Elsevier page "Article Sharing" says, *"Theses and dissertations which contain embedded PJAs [Published Journal Articles] as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect"* [23]. The references [6, 7] in Chapter 8 include DOIs, so I believe this requirement has been met.

Acknowledgements

I would like to thank Josie, Stephanie, Christabel and Max for much support and patience while I completed this work.

I would like to thank my supervisors, Prof. Nigel Bean and Prof. Joshua Ross, for continual guidance and countless meetings.

I would like to thank the University of Adelaide for the resources it has provided to make this work possible.

1 Introduction

1.1 Epidemic fade-out

"Fade-out" refers to the situation in which an infectious disease fades out in a population. That is, it reaches the point where no individuals are infectious or carry the infection. Since this is in the context of an infectious disease, this means that the disease is eliminated from the population.

Two ways in which an infection might fade out are reasonably intuitive. A disease may be introduced into a population, but then fade out in the initial stage before taking hold, failing to infect a significant number of individuals. There does not appear to be a standard term for this, so we shall refer to it as *"initial fade-out"*. This is illustrated in Figure 1, where initial fade-out occurs if the infection fades out near point A.

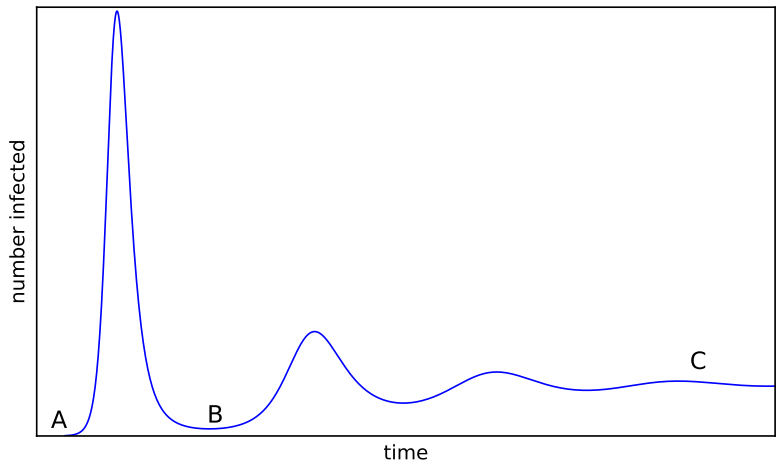


Figure 1: A possible plot of number infected, versus time. An infection may fade out near point A (initial fade-out), or become endemic (point C). Epidemic fade-out refers to fade out during the trough after an initial wave of infection, near point B.

Alternatively, an infection might persist through the initial stage, and become endemic in the population, as illustrated by an infection reaching a point

near C in Figure 1. It may then later fade out due to a change in circumstances, or simply due to random effects. In this case, the mean time to fade-out will usually be quite long, and it is referred to as *"endemic fade-out"* [2].

"Epidemic fade-out", the subject of this thesis, is distinct from either of these [2], and refers to the situation when an infection fades out in the first trough after an initial major wave of infection. This is illustrated by fade-out occurring near point B in Figure 1. In contrast to an initial fade-out, there is a major outbreak and a significant number of individuals are infected in that outbreak. In contrast to endemic fade-out, epidemic fade-out occurs a relatively short time after the initial outbreak.

A condition for epidemic fade-out is that the outbreak contains oscillations. If one plots the number of infected individuals against time, then oscillations may occur [2]. When the oscillations reach their low point, stochastic effects may cause the number of infected individuals to fall to zero, effecting fade-out. In some situations, the first peak is particularly high, and the first trough is particularly low, as in Figure 1. So fade-out during this first trough is more likely than during the second and subsequent troughs, and is worthy of separate consideration [19]. Epidemic fade-out generally only refers to fade-out in this first trough, and that is the usage assumed in this thesis.

The term "epidemic fade-out", though used by many authors [2, 9, 20, 48, 59] (or without the hyphen: "epidemic fadeout" [43]) is not universal. Other terms which have been used are "extinction in the first trough" [18], "fade-out post epidemic" [21], and "extinction at the end of a major outbreak" [59]. These all refer to the same phenomenon.

1.2 Motivation and aims

The motivation for this work is that a number of prominent researchers mention epidemic fade-out as a phenomenon worthy of further research, in order to gain a better understanding of the dynamics of outbreaks of an infectious disease.

The first attempt to calculate the probability of epidemic fade-out was by van Herwaarden in 1997 [59], which will be discussed in more detail in Section 2.4.3. Commenting on this paper, Diekmann and Heeterbeek wrote in 2000, *"In fact we only know one paper in which the relevant probability is calculated... It is hoped that this will trigger more work in this direction, concentrating on other models and different methods such that in the end a more robust picture emerges"*[22].

Nevertheless, in the intervening years, only one further paper emerged, by Meerson and Sasorov in 2009 [48]. This again led to suggestions that more work should be done in this area. In a 2013 seminar, Britton said that the problem was *"Not well solved even for simplest model!"* (sic) [18]. Then a 2015 paper co-authored by Britton listed further understanding of epidemic fade-out as one of "Five challenges for stochastic epidemic models involving global transmission" [19]. They wrote, *"a more challenging question is how to calculate the probability that the infection persists through the trough that follows the initial epidemic"*. After noting the published work of van Herwaarden and Meerson and Sasorov, they wrote *"Challenges remain in extending this work beyond the simplest settings"*.

Therefore this work was undertaken to extend the understanding of epidemic fade-out.

The work contains two main themes, spread over two published papers and one submitted paper. The first theme is the *calculation* of the probability of epidemic fade-out. This is the focus of the first [6] and third [8] papers. In

addition to calculating this probability, these papers examine, and in some cases explain, the factors which affect this probability.

The other theme, and the topic of the second paper [7], is the *control* of epidemic fade-out. The first paper included the somewhat surprising result that the probability of epidemic fade-out has a non-monotonic relationship with β, the transmission rate parameter. Therefore this paper considers strategies for controlling β in order to maximise the probability of epidemic fade-out. In particular, it provides a simple formula for a close-to-optimal strategy.

All of the work in this thesis uses the Markovian SIR-with-demography model [51]. In addition, the third paper extends the work to also apply to the Markovian SIRS model [30], which is closely related.

1.3 Organisation of this thesis

This work is a "thesis by publication". During the course of my candidature I submitted three papers in which I was the lead author. This thesis refers to these as Paper 1, Paper 2 and Paper 3. Papers 1 [6] and 2 [7] were published in 2016 and 2017 respectively, while Paper 3 [8], the most recent, is still under review. These papers form the main body of the thesis.

Chapter 2 includes an in-depth description of the SIR-with-demography and SIRS infection models, and reviews the literature; both of these infection models, and of work investigating epidemic fade-out.

Chapters 3, 4 and 5 contain the publications themselves. Each chapter includes a short introduction to the paper, a "Statement of Authorship" declaration, and the paper itself. Chapters 3 and 4 contain reprints of the published versions of Papers 1 and 2, respectively. Chapter 5 contains the submitted version of Paper 3.

Chapter 6 explains the reasons for the choices of software used.

The conclusion in Chapter 7 summarises the results, and considers possible directions for future research.

2 Model and Literature review

2.1 The SIR-with-demography model

The main model used in this thesis is the "SIR-with-demography model" [51], although Paper 3 [8] also uses the "SIRS model" [30].

The use of a model allows assumptions to be made. Almost any model will significantly simplify the physical world, but a well-chosen model will closely approximate the physical world, and allow observations and predictions to be made about real world systems. The Markovian SIR model has a long history and is well tested. The use of demography – adding births and deaths to an otherwise static population – allows the modelling of changes in population, and also has a long history of usage. The SIRS model is a closely related model, which models waning immunity of an infection.

But the SIR-with-demography and SIRS models are also simple enough to compute results reasonably effectively. Therefore we concentrate on these models. The SIR-with-demography model also has the advantage that it is the model which was used in previous studies of epidemic fade-out [59, 48], allowing us to compare our results with theirs.

The SIR-with-demography model and the SIRS model are both *compartmental models*. In a compartmental model, every individual in the population is put into one of a number of compartments. Individuals in the same compartment are assumed to all have the same characteristics. Therefore, to completely describe the system, all that is required is to specify the number of individuals in each compartment.

Both of these models are variants of the basic "SIR infection model". This is used to model the spread of a disease with two main properties: first, the disease is *infectious*, and is spread by the contact between an infectious individ-

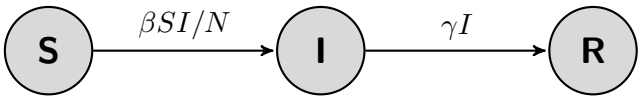


Figure 2: The SIR infection model

ual and a susceptible individual; and second, recovery from the disease grants *immunity*, so that those who have recovered are permanently immune. (In the SIRS model, which is discussed in Section 2.3, immunity is not permanent; and wanes with time.)

The SIR infection model takes its name because it consists of three compartments, "S", "I" and "R"; and these three compartments are also used in the SIR-with-demography and SIRS models.

"S" stands for "susceptible", and individuals in this compartment are susceptible to the disease in question.

"I" stands for "infectious", and individuals in this compartment are infectious, capable of spreading the disease.

"R" stands for "recovered" (or according to some authors, "removed"). Individuals in this compartment have recovered from the infection and, importantly, are immune from further infection. They play no part in the further spread of the disease, which is why they are sometimes treated as removed from the population.

As a convention, this thesis will use quote marks to denote the compartment itself, and the same letter without quotes to denote the number of individuals in that compartment. So the "S" compartment contains S individuals, the "I" compartment contains I individuals, and the "R" compartment contains R individuals.

The transition rates between states in the SIR model are shown in Figure 2. The model assumes that infections occur at a rate proportional to both S and I. Each infectious individual has potentially infectious contacts with other

individuals, at an average rate β. However, only the susceptible individuals can be infected, so the probability of the contact actually causing an infection is approximately S/N, where N is the population size. This gives an infection rate of $\beta SI/N$.

For small population sizes, an infection rate of $\beta SI/(N-1)$ is more correct, because an individual cannot contact or infect themselves. However in larger populations, as considered in the SIR-with-demography model used in this thesis, $N \approx N - 1$, so the difference is trivial.

Recovery from infection is assumed to occur at rate γ (that is, the infectious time is exponentially distributed with mean infectious time $1/\gamma$). This gives rise to a net rate of γI from the "I" state to the "R" state.

The SIR model, and extensions of it considered in this thesis, assume that the populations can be treated as homogeneous – that is, that all individuals have the same infection and recovery rates, or at least that it is a valid working assumption that average rates can be used. It also assumes the population is well mixed, so that all individuals have an equal probability of making an infectious contact. Both of these assumptions can be relaxed by making the model more complicated, but that is outside the scope of this thesis.

The SIR-with-demography model extends the SIR model by adding demography, and is shown in Figure 3. These might be birth and death rates, or immigration and emigration rates. For convenience, we shall refer to them as birth and death rates in this thesis, but immigration and emigration is not excluded. Per capita deaths occur from each state at rate μ (corresponding to an exponentially distributed lifetime with mean $1/\mu$), and births occur at rate μN.

As this is a stochastic model, the population size is not fixed, but varies with time. So the meaning of N is that it is the mean equilibrium population size [3].

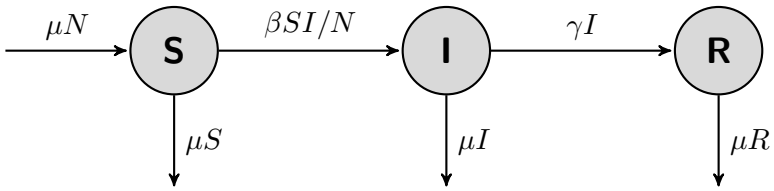


Figure 3: The SIR-with-demography epidemic model.

(That paper adds the qualification "when the infection has been eradicated", but this qualifier is not necessary, because the death rate for individuals is the same, regardless of what state they are in). There is also no upper limit on the population size, so the state space is unbounded, an issue which we had to address in this research.

Note that R does not figure in any of the rate equations for the "S" or "I" states. Therefore it is possible to remove the "R" state. This gives the representation of the SIR-with-demography model used in this research, as shown in Figure 4. The corresponding transition rates are shown in Table 1. An important feature is that the state space is two-dimensional, since there are only two variables, S and I.

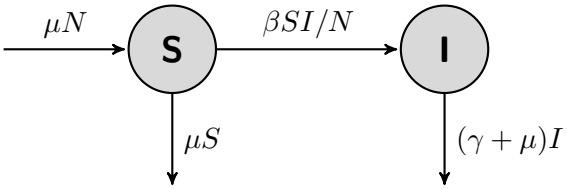


Figure 4: The SIR-with-demography epidemic model, without the "R" state, as used throughout this thesis.

Inevitably, not every author uses the same symbols when describing their work mathematically. In this thesis, unless indicated otherwise, we use the terminology in Figure 4 and Table 1 when discussing other researchers' work. We also use p_0 as shorthand for "probability of epidemic fade-out". However not all authors have an identical definition of p_0, and one issue we encountered

Description	Transition	Rate
Infection	$(S, I) \to (S-1, I+1)$	$\beta SI/N$
Birth of Susceptible	$(S, I) \to (S+1, I)$	μN
Death of Susceptible	$(S, I) \to (S-1, I)$	μS
Removal of Infectious	$(S, I) \to (S, I-1)$	$(\gamma+\mu)I$

Table 1: Transition rates for the Markovian SIR-with-demography epidemic model displayed in Figure 4.

was that no one else had attempted to define p_0 precisely.

Important concepts, which appear throughout this thesis, are the *deterministic approximation* and *endemic point*. In the deterministic approximation (also known as the deterministic model [59]), instead of the stochastic rate equations in Table 1, the system is specified by the differential equations (where t is time):

$$\begin{aligned} dS/dt &= \mu(N-S) - \beta SI/N, \\ dI/dt &= \beta SI/N - (\gamma+\mu)I. \end{aligned} \tag{1}$$

In the deterministic approximation, assuming a starting condition with positive I, the system converges to a stable equilibrium point, which is sometimes called the endemic point [39]. This point has the values $S = S_e$ and $I = I_e$, and is the point at which $dS/dt = 0$ and $dI/dt = 0$. It is not difficult to show that,

$$(S_e, I_e) = N\left(\frac{\gamma+\mu}{\beta}, \frac{\mu(\beta-\gamma-\mu)}{\beta(\gamma+\mu)}\right). \tag{2}$$

2.2 Historical development of the model

The SIR infection model has a long history. The concept of "mass action" – of the rate of infection being proportional to both the number of susceptible

individuals and the number of infectious individuals – appears to originate with Hamer in 1906 [1].

But the SIR model itself is generally credited to originate with Kermack and McKendrick's 1927 paper [36], with the SIR model sometimes called the "Kermack-McKendrick model" [1].

Kermack and McKendrick introduced a *deterministic* model. That is, the rates in Figure 2 were deterministic rates; so the model can be represented by the differential equations:

$$\begin{aligned} dS/dt &= -\beta SI/N \\ dI/dt &= \beta SI/N - \gamma I. \end{aligned} \tag{3}$$

They demonstrated the efficacy of the model by showing how it matched the data from an outbreak of plague in 1905, as shown in Figure 5.

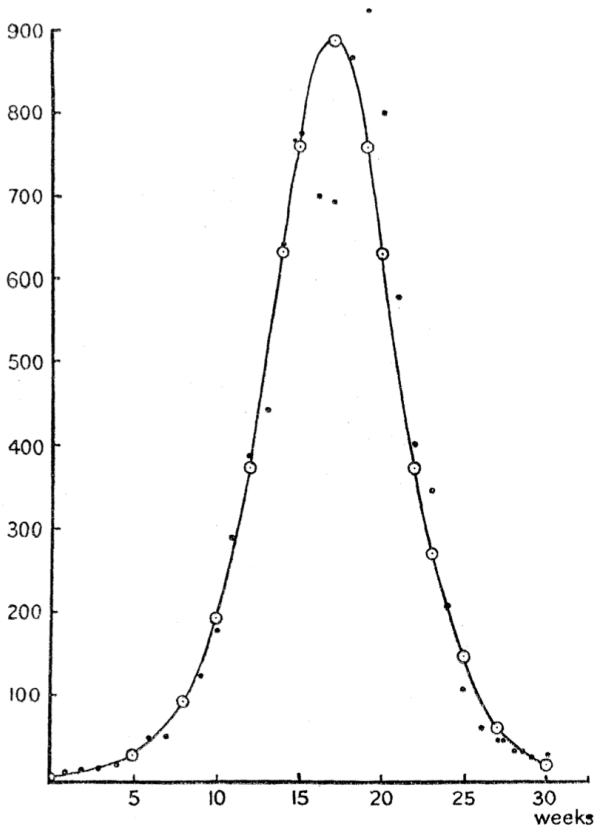
However the SIR model has no way of supplying new susceptibles (dS/dt is never positive in (3)). The addition of demography (to give the SIR-with-demography model) is usually credited [1] to Soper's 1929 paper [56]; and Bartlett in 1956 refered to it as "the Hamer-Soper model" [11]. However it appears more accurate to credit it to an earlier (1921) paper by Martini [44], which has the equations in a slightly different form. Therefore Nåsell refers to the SIR-with-demography equations as "the Martini model" [50].

The SIR-with-demography model has become very widespread, with some papers even referring to it as the "basic SIR model" [43] or the "Classic Endemic Model" [31]. This indicates how ubiquitous it is, and why it is a good choice of model to study.

Like Kermack and McKendrick, both Martini and Soper used deterministic models. However deterministic models cannot account for fade-out; stochastic models are required for this.

Also for the rate at which cases are removed by death or recovery which is the form in which many statistics are given

$$\frac{dz}{dt} = \frac{l^3}{2x_0\kappa^2}\sqrt{-q}\,\mathrm{sech}^2\left(\frac{\sqrt{-q}}{2}lt - \phi\right). \qquad (31)$$



The accompanying chart is based upon figures of deaths from plague in the island of Bombay over the period December 17, 1905, to July 21, 1906. The ordinate represents the number of deaths per week, and the abscissa denotes the time in weeks. As at least 80 to 90 per cent. of the cases reported terminate fatally, the ordinate may be taken as approximately representing dz/dt as a function of t. The calculated curve is drawn from the formula

$$\frac{dz}{dt} = 890\,\mathrm{sech}^2\,(0{\cdot}2t - 3{\cdot}4).$$

Figure 5: The figure from Kermack and McKendrick's 1927 paper, which illustrates the accuracy of the deterministic SIR infection model

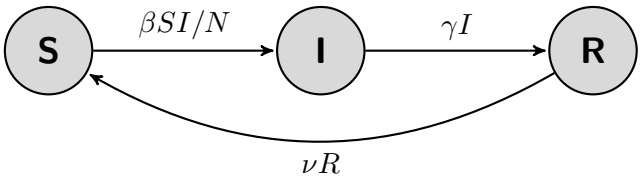


Figure 6: The SIRS model

McKendrick in 1926 [47] and Greenwood in 1931 [28] attempted stochastic treatments of infections, but not with SIR-type models.

Stochastic treatments of the SIR-related models emerged in the late 1940s and 1950s. The origin of stochastic analysis of the SIR model was touched upon by Bartlett in 1949 [10], then Bailey in 1950 published a more thorough stochastic treatment of the SIR model (without demography) in 1950 [4]. The first stochastic treatment of the SIR-with-demography model appears to have been by Bartlett in 1956 [11].

So we see, from 1956 onwards, the stochastic SIR-with-demography model regularly appearing in the literature, and this is the main model used in this thesis.

2.3 The SIRS model

The SIR-with-demography model is the main model used in this work, and the only model used in Papers 1 and 2 [6, 7]. However, in Paper 3 [8], we also use the closely related SIRS ("Susceptible-Infectious-Recovered-Susceptible") model. This models the renewal of susceptibles via waning immunity (that is, infectious individuals who are already in the community, can move from the "R" state to the "S" state); rather than introducing new individuals to the community. This is a more appropriate model to use for diseases with an immune time which is much shorter than individuals' lifetime, such as

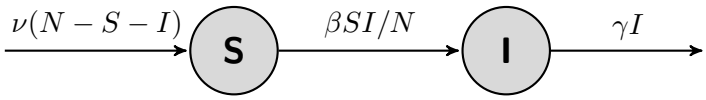


Figure 7: The SIRS model with the "R" state removed

Description	Transition	Rate
Infection	$(S, I) \to (S-1, I+1)$	$\beta SI/N$
Removal of Infectious	$(S, I) \to (S, I-1)$	γI
Loss of Immunity	$(S, I) \to (S+1, I)$	$\nu(N-S-I)$

Table 2: Transition rates for the Markovian SIRS epidemic model displayed in Figure 7.

influenza. The SIRS model appears to have been introduced by Hethcote in 1976 [30].

The SIRS model is illustrated in Figure 6. A property of the SIRS model is that the population is constant, with N being the fixed population size. Therefore $R = N - S - I$ and the "R" state can be removed, as in Figure 7, with the transition rates shown in Table 2

For our purposes, there is no qualitative difference in behaviour between the two models. In the software, they are both modelled with a single model, illustrated in Figure 8, with the condition that exactly one of μ and ν must be zero.

A combination in which μ and ν are both non-zero is the SIRS-with-demography model [24, 51], and represents a system in which both births

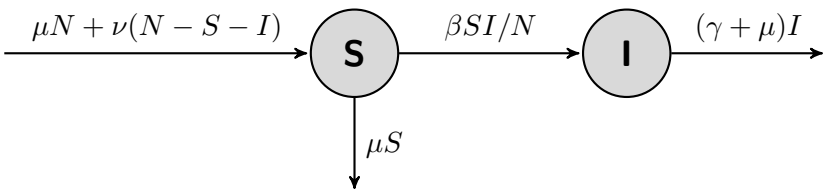


Figure 8: Combined model of SIR-with-demography and SIRS, in which exactly one of μ and ν must be zero.

and deaths, and waning immunity, are modelled. However (unlike the SIRS model), the population size is not fixed, so the assumption $R = N - S - I$ no longer holds; and (unlike the SIR-with-demography model), R *does* feature in a rate equation for S. Therefore, unlike the SIRS and SIR-with-demography models, the "R" state cannot be removed. This means the model has a three-dimensional state space – compared to a two-dimensional state space for the SIRS and SIR-with-demography models. Since we wish to keep the models simple, this put the SIRS-with-demography model outside the scope of this thesis.

2.4 Epidemic fade-out and related problems

In terms of issues related to fade-out, Ball in 1983 largely solved the problem for initial fade-out [5] – that is, fade-out before a major outbreak occurs. A specific treatment for the SIR-with-demography model was given in 1995 [60].

In terms of endemic fade-out, the question is one of the time for fade-out to occur. Numerous works have examined the Mean Time to Extinction (MTE) in a general stochastic case [35, 41], in related models such as the logistic model [49, 54] or SIRS with demography [51], and in the specific case of SIR-with-demography [60, 50].

2.4.1 Critical Community Size (CCS)

In comparison to initial and endemic fade-out, epidemic fade-out has been relatively neglected. The term, and its distinguishment from endemic fade-out, appears to originate with Anderson and May [2].

However, the study of CCS (Critical Community Size) is in many ways related. CCS was proposed by Bartlett in relation to measles. He determined that the CCS for measles (in the pre-vaccination era; vaccination began in the

1960s in most developed countries [2, 23, 31]) is between about $200,000$ and $300,000$ [12, 13]. This study is regarded as a classic; both the paper and its result is widely cited, albeit at a slightly revised figure of $250,000$ to $500,000$ [2, 29, 21].

The reason for relating CCS to epidemic fade-out is that CCS implicitly refers to the situation of an outbreak fading out soon after its major wave of infection. We are aware of only one paper [43] which connects the terms CCS and epidemic fade-out directly, although a number of studies connect the concepts [29, 22, 57, 52, 63].

It has been pointed out that below the CCS, an infection dies out "in the troughs between epidemics" [29], which is very much the definition of epidemic fade-out.

A number of papers point out that epidemic fade-out is caused by the pool of susceptibles running out, due to the intensity of the infection [57, 43, 63]. Nåsell is somewhat more precise, saying, *"We interpret this to say that the critical community size is that value of N for which the probability of extinction after waiting for one quasi-period T_0 equals 0.5"* [52], where the quasi-period T_0 is the period of oscillations in the number of infectious individuals. That is, Nåsell explicitly says that CCS is tied to fade-out within a fixed period of time. However, none of these papers make an attempt to directly calculate the probability of epidemic fade-out.

Diekmann and Heesterbeek make some heuristic obervations about CCS in the SIR-with-demography model, and even estimate the relative influence of μ and N, with the CCS roughly depending on $N\sqrt{\mu}$ relative to other parameters. They say that a higher $N\sqrt{\mu}$ makes a recurrence of outbreaks more likely (that is, $p_0 \approx 1$), and that a lower $N\sqrt{\mu}$ makes a single outbreak more likely (that

is, $p_0 \approx 0$)[1]. For other situations they simply say "everything else in between could be called critical" [22].

Swinton *et al.* [57] divide SIR-with-demography outbreaks into three types: those with low birth rates (that is, low μ), which always fade out after the first wave; those with high birth rates, which never fade out after the first wave; and those which are intermediate, which may or may not exhibit epidemic fade-out, depending on stochastic effects. They note that these types are affected by μ (epidemic fade-out becomes less likely as μ increases), and by N (epidemic fade-out becomes less likely as N increases), but make no attempt to quantify these factors.

Similarly, Lloyd-Smith *et al.* [43] make a number of useful qualitative observations about epidemic fade-out, but go no further. Like Swinton *et al.*, they note the dependence of the probability of epidemic fade-out on N (implying that the probability of epidemic fade-out falls as N increases). They also note that there is no "abrupt" change in this probability in most cases – in other words, there is no clear-cut CCS at which there is a dramatic drop in the probability of epidemic fade-out, instead noting that epidemic fade-out depends on a number of factors which "depend on N in complex ways". In this regard, they posit that pre-vaccination measles, which does exhibit a reasonably clear CCS, is the exception rather than the rule. As with Swinton *et al.*, they point out that epidemic fade-out depends very much on the value of μ. They also point out that epidemic fade-out depends on (in their words) the "intensity" of the epidemic (because a more intense epidemic depletes the population of more susceptibles). But they do not quantify this "intensity", nor do they attempt to calculate the probability of epidemic fade-out.

Xiao *et al.* [63] perform many random simulations and infer a number of

[1]In fact there is a slight mistake in their written text – on page 48 they write "small' where they mean "large", and vice versa – but their formulae are correct.

conclusions from them. But they too make no attempt at a formula for the probability of epidemic fade-out, or for CCS, from the parameters.

Similarly to Diekmann and Heesterbeek, Nåsell makes a number of approximations and derives some simplified expressions for the CCS in the SIR-with-demography model [52].

So while there is a lot of work on CCS, it is either inferred from observations, or based on some simplifying assumptions. While they contain some very useful information, none of these are directly useful in calculating the probability of epidemic fade-out.

2.4.2 Probability of epidemic fade-out – introduction

On the specific subject of epidemic fade-out, only two papers existed before ours, as previously mentioned in Section 1.2. We now discuss these in greater detail.

Both papers consider the probability of epidemic fade-out for the SIR-with-demography model. We are not aware of any attempts to calculate the probability of epidemic fade-out in any other model.

2.4.3 van Herwaarden, 1997

van Herwaarden [59], was the first person to calculate an approximation for the probability of epidemic fade-out, given a set of parameters. He uses the SIR-with-demography model as specified in Figure 4 and Table 1.

van Herwaarden makes the point that if the renewal rate (μ) is very high, then the probability of epidemic fade-out is close to 0; and if μ is very low, then the probability of epidemic fade-out is close to 1. He writes that his derivation specifically pertains to the intermediate case. (This distinction, between three different types of outbreaks, was later used by Swinton *et al.* in their discussion

of CCS [57], which was mentioned above).

van Herwaarden first calculates expressions for the path of the outbreak in the deterministic model. That is, he calculates the evolution in time of (S, I) given the differential equations (1),

$$dS/dt = \mu(N - S) - \beta SI/N$$
$$dI/dt = \beta SI/N - (\gamma + \mu)I;$$

and an initial condition of small I, and $S = N - I \approx N$. The deterministic path is shown in Figure 9, which is copied from van Herwaarden's paper.

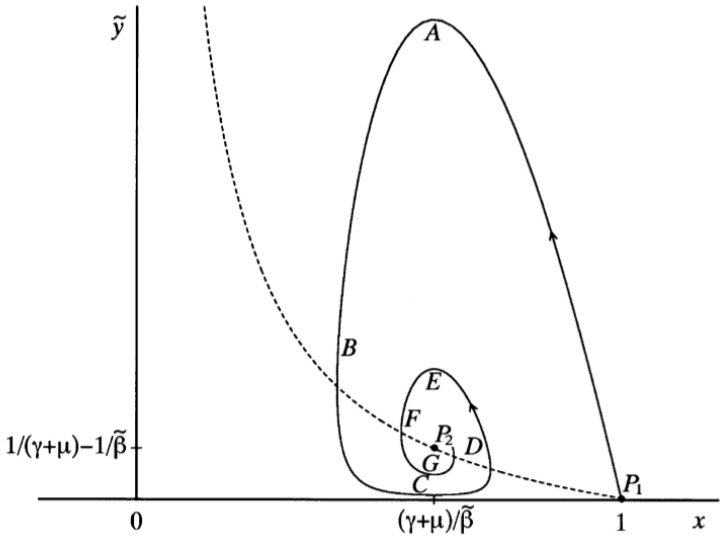


Figure 9: [59, Figure 2], showing the deterministic path of an outbreak from an initial point P_1, using $x = S/N$, $\tilde{y} = I/(N\mu)$ and $\tilde{\beta} = \beta$. The path is divided into segments A through to G. The dotted line shows points where the tangent to the deterministic path is vertical. P_2 is the endemic equilibrium point.

van Herwaarden breaks the path into segments, A through to G. For each segment he derives an expression for the path, using a local asymptotic expansion. He derives one expansion for segments A and E (where I is high), another expansion for segments C and G (where I is low), and a third expan-

sion for segments B, D and F (where I is varying most dramatically). He then matches adjacent expansions at their boundaries. The result is that he obtains a single expression for the deterministic trajectory up to and including segment C, the region in which epidemic fade-out may occur.

In effect, van Herwaarden breaks the problem into two components. For the first part of the problem (segments A and B) he assumes that a deterministic solution is sufficient. He does not elaborate on the reason for this (simply saying "This part of the process is dominated by the deterministic field"), but implicitly the reason is that both S and I are sufficiently large for the deterministic approximation to hold to a good degree of accuracy.

van Hewaarden then sets up a two boundary problem in segment C: an absorbing boundary at $I = 0$, and an artificial upper absorbing boundary at $I = I_e$, where I_e is the endemic I value (2). To solve this problem – that is, to calculate the probability of first hitting a particular absorbing boundary – he uses the Fokker-Planck equation [25], and calculates an approximate solution.

Then, using a point from the deterministic analysis as the starting condition, he obtains an approximate expression for the probability of epidemic fade-out. The expression, which is explicit but rather complicated, is given in our Paper 1 [6, Equation (8)].

This expression assumes an initially naive population, (that is, P_1 in Figure 9 is at $x = 1$ and $\tilde{y} = 0$, corresponding to $S = N$ and $I \approx 0$, though of course in reality the initial I must be non-zero). The final section of van Herwaarden's paper shows how to use the same techniques to calculate the probability of epidemic fade-out for different initial conditions.

So in summary, van Herwaarden derives an explicit expression which is approximately equal to the probability of epidemic fade-out. There are limitations – he notes that his asymptotic approximations of the Fokker-Planck

equation assumes a large[2] N – but as we shall see in Paper 1 [6], the results are still quite accurate, though not as accurate as the numerical method we devise in that paper.

2.4.4 Meerson and Sasorov, 2009

The only other previous paper to attempt to calculate the probability of epidemic fade-out, was by Meerson and Sasorov in 2009 [48]; though some of the details were discussed in a subsequent paper which Meerson co-authored [55].

Like van Herwaarden, they set up a two boundary problem. Due to their choice of parameters, the artificial upper absorbing boundary was slightly different, a diagonal line from (S_e, I_e) to $(N, 0)$ instead of a horizontal line at $I = I_e$ (where S_e and I_e are as specified in (2)). The simulations and calculations we conducted suggest that this choice of upper absorbing boundary made little difference: if a realisation reached this boundary, it would almost invariably also reach the boundary defined by van Herwaarden.

They borrow van Herwaarden's deterministic analysis, and use his expressions to give the same starting point for their two boundary problem. But instead of using the Fokker-Planck equation, they use the WKB approximation [15],

The WKB approximation is named after three theoretical physicists – Wentzel, Kramers and Brillouin – who all published works using it in 1926 [62, 38, 17]. It is sometimes alternatively called the Liouville-Green or LG approximation [26], because it was later realised [53] that the method was discovered separately by Liouville [42] and Green [27] in 1837.

The WKB approximation improves on a shortcoming in the Fokker-Planck approximation, that the latter very often fails to model large fluctuations cor-

[2]He writes $1/\sqrt{N} \ll \mu \ll 1$, but this cannot be correct because μ is not dimensionless. It appears that $1/\sqrt{N} \ll \mu/\gamma \ll 1$ is intended.

rectly [48]. This can make it unsuitable in applications where large fluctuations are the main cause of fade-out [55]. The WKB approximation is an expansion method which more accurately predicts these large fluctuations [55].

In the context of the probability of epidemic fade-out, the method is only claimed to hold for N large enough that the probability of epidemic fade-out is near zero; though we observed it was quite accurate across a range of values [6]. Like van Herwaarden, they derive an expression which is explicit but rather complicated. The expression is given in Paper 1 [6, Equation (9)].

Nevertheless, we did not observe a significant increase in accuracy (over the use of the Fokker-Planck method as used by van Herwaarden [59]) in their probability of epidemic fade-out calculations [6]. One possible reason is the mechanics of epidemic fade-out: the WKB approximation was claimed to be more accurate for large fluctuations, and large fluctuations are the major cause of *endemic* fade-out, which was the subject of earlier papers which applied the WKB approximation to ecological modelling [37, 34]. It may be that, compared to endemic fade-out, epidemic fade-out is less reliant on large scale fluctuations. Simulations during the work for this thesis indicated that epidemic fade-out most commonly occurred when the deterministic path of the outbreak took the number of infectious individuals (I) to quite a low value; and from that point only a relatively small fluctuation is required for fade-out to occur. In terms of numbers of realisations (and hence, in terms of the overall contribution to probability of epidemic fade-out), this mode of fade-out far exceeded those due to large fluctuations. That may explain why the WKB approximation's superiority in modelling large fluctuations was not a noticeable factor.

It may also mean that the WKB method is in fact superior when epidemic fade-out is a rare event (that is, when p_0 is very low). However we only compared the methods for situations when epidemic fade-out was reasonably

probable, choosing parameter sets for which p_0 was roughly between 0.1 and 0.9 [6].

3 Paper 1

3.1 Introduction

Paper 1 is entitled "The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography". It was published in *Journal of Theoretical Biology* in 2016 [6].

As we observed in the literature review, two very good approximate formulae for p_0 had already been published [59, 48]. However, these were still sufficiently complex that it was difficult to make qualitative conclusions about the phenomenon of epidemic fade-out.

Therefore we devised a numerical calculation method, that we showed was more accurate (but slower to compute) than the previously published formulae, and faster (though sometimes less accurate) than more exact methods.

Then in the latter part of the paper, we used this approximation to survey trends in p_0, and observed that it was non-monotonic in β, the transmission rate parameter. We also offered an explanation for why this occurs.

We concluded by suggesting possible applications of this non-monotonicity, as well as pointing to the need to examine epidemic fade-out in other models. These became the directions for Papers 2 and 3.

3.2 Statement of Authorship

Statement of Authorship

Title of Paper	The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography
Publication Status	**Published**
Publication Details	P. G. Ballard, N. G. Bean, J. V. Ross, The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography, Journal of Theoretical Biology, 393:170-178, 2016, doi:10.1016/j.jtbi.2016.01.012

Principal Author

Name of Principal Author (Candidate)	Peter Ballard		
Contribution to the Paper	Derived most of the mathematics, wrote all the algorithms and code, ran all the code, generated all the diagrams, wrote most of the text.		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20-3-2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Prof. Nigel Bean		
Contribution to the Paper	Project supervision, idea generation, helped analyse the data, some specific mathematical suggestions, suggestions and corrections to text.		
Overall percentage (%)	10%		
Signature		Date	28/03/2018

Name of Co-Author	Prof. Joshua Ross		
Contribution to the Paper	Initial concepion of project, project supervision, idea generation, helped analyse the data, suggestions and corrections to text, wrote first draft of the introduction.		
Overall percentage (%)	10%		
Signature		Date	28/03/18

3.3 Paper 1

The paper, as published in *Journal of Theoretical Biology* in 2016, is on the following pages.

Journal of Theoretical Biology 393 (2016) 170–178

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography

P.G. Ballard [a,*], N.G. Bean [a,b], J.V. Ross [a]

[a] School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia
[b] ARC Centre of Excellence for Mathematical and Statistical Frontiers, Australia

HIGHLIGHTS

- We calculate the probability of epidemic fade-out for the SIR model with demography.
- We present an efficient algorithm, more accurate than published approximations.
- The probability of epidemic fade-out generally peaks near $R_0 = 2$.
- We explain why epidemic fade-out is more likely near $R_0 = 2$.

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form
22 December 2015
Accepted 7 January 2016
Available online 18 January 2016

Keywords:
Diffusion approximation
Efficient algorithms
Epidemic control
Stochastic epidemic model

ABSTRACT

Epidemic fade-out refers to infection elimination in the trough between the first and second waves of an outbreak. The number of infectious individuals drops to a relatively low level between these waves of infection, and if elimination does not occur at this stage, then the disease is likely to become endemic. For this reason, it appears to be an ideal target for control efforts. Despite this obvious public health importance, the probability of epidemic fade-out is not well understood. Here we present new algorithms for approximating the probability of epidemic fade-out for the Markovian SIR model with demography. These algorithms are more accurate than previously published formulae, and one of them scales well to large population sizes. This method allows us to investigate the probability of epidemic fade-out as a function of the effective transmission rate, recovery rate, population turnover rate, and population size. We identify an interesting feature: the probability of epidemic fade-out is very often greatest when the basic reproduction number, R_0, is approximately 2 (restricting consideration to cases where a major outbreak is possible, i.e., $R_0 > 1$). The public health implication is that there may be instances where a non-lethal infection should be allowed to spread, or antiviral usage should be moderated, to maximise the chance of the infection being eliminated before it becomes endemic.

1. Introduction

The ultimate goal of modelling infectious disease dynamics is to gain insight into how to use resources best to eliminate infection. This may be achieved by making invasion difficult through minimising the probability of a major outbreak, for example through the use of prophylactic vaccination, antivirals or contact tracing (Ball, 1983; Ball and Lyne, 2002; Ross and Black, 2015).

For endemic diseases, with wide prevalence, once again the predominant focus is on reducing transmission as much as possible, and there have been a number of studies calculating the mean time to *endemic fade-out* (van Herwaarden and Grasman, 1995; Nåsell, 2001; Kamenev and Meerson, 2008).

Much less attention has been paid to what is the optimal approach to adopt when a major outbreak occurs. Typically, focus has been given to minimising the amount of infection – either the rate of new infections, or the total number of infections over the first wave of an outbreak – for example, through the use of antivirals, and once available, vaccination (e.g., McCaw and McVernon, 2007; Black et al., 2013). Here we instead focus on the probability of *epidemic fade-out* – that is, the probability of infection being eliminated between the first and second waves of infection.

In fact, a more comprehensive understanding of the probability of epidemic fade-out is named as one of the five challenges (for stochastic epidemic models involving global transmission) by Britton et al. (2015), supporting earlier calls (Anderson and May, 1991; Diekmann and Heesterbeek, 2000). The interest in this

* Corresponding author.
E-mail addresses: peter.ballard@adelaide.edu.au (P.G. Ballard), nigel.bean@adelaide.edu.au (N.G. Bean), joshua.ross@adelaide.edu.au (J.V. Ross).

http://dx.doi.org/10.1016/j.jtbi.2016.01.012
0022-5193/

quantity for infection elimination is that following the first wave of an outbreak, the number of infectious individuals drops to a relatively low level. Then, if fade-out does not occur, it is likely that the disease will become endemic. Hence, this "*first trough*" of infection appears intuitively to be an ideal target for elimination.

We study a Markovian SIR model with demography (van Herwaarden and Grasman, 1995; Nåsell, 1999; Andersson and Britton, 2000), and in particular the probability of epidemic fade-out as a function of effective transmission rate, recovery rate, population turnover rate, and population size parameter. We identify the ubiquity of a non-monotonicity property of the probability of epidemic fade-out as a function of effective transmission rate (holding other parameters fixed). In fact, the probability of epidemic fade-out is very often greatest when the basic reproduction number, R_0, is approximately 2 (restricting consideration to cases where a major outbreak is possible, i.e., $R_0 > 1$). This means that there may be cases when, faced with an infectious outbreak, it would be beneficial to not take action to reduce R_0.

The identification of this phenomenon was achieved through the development of a numerical method which is highly accurate and efficient for computation of the probability of epidemic fade-out. To our knowledge, as supported by the paper (Britton et al., 2015), there have been only two existing methods proposed, both approximations, for evaluating this probability (van Herwaarden, 1997; Meerson and Sasorov, 2009). These existing methods are asymptotic approximations, with accuracy improving in the limit as the population size parameter tends to infinity. Our method has the benefit of being highly accurate across a wider range of population sizes, including moderate-sized populations, while still using light computer resources and hence scaling well to large population sizes.

In the next section we introduce the Markovian SIR model with demography that we study, before discussing deterministic and diffusion approximations of this model which are relevant to existing methods and our new method for evaluating the probability of epidemic fade-out. We then review the existing approximations. In Section 3 we detail our new method for computing the probability of epidemic fade-out. In Section 4.1 we validate its accuracy and efficiency, and in Section 4.2 we investigate the dependence of the probability of epidemic fade-out on the model parameters, identifying the ubiquity of a non-monotonicity property in the effective transmission rate. Finally, we conclude this work and discuss future research ideas.

2. Background

In this section we present the two existing methods for approximating the probability of epidemic fade-out (van Herwaarden, 1997; Meerson and Sasorov, 2009). To achieve this, we first introduce the underlying model assumed in these earlier studies, and also two asymptotic approximations of this model. These are not only required for both existing methods but also for our new methods to be presented in Section 3.

2.1. The Markovian SIR model with demography

Following previous work (van Herwaarden, 1997; Meerson and Sasorov, 2009), we adopt the Markovian SIR model with demography (van Herwaarden and Grasman, 1995; Nåsell, 1999; Andersson and Britton, 2000). However, we note that our methods can be easily modified to suit other SIR models which involve replenishment of susceptibles.

The well-known SIR model puts every individual in the population into one of three classes: "S" for Susceptible, "I" for Infectious, and "R" for Recovered (or Removed). Let S, I and R denote the number of individuals in the respective states. Then, we assume that susceptible individuals become infectious at rate $\beta SI/N$, and infectious individuals recover at rate γI, where β is the effective transmission rate parameter, $1/\gamma$ is the average infectious period of an individual and N is the total population size. The population is closed, and hence of a constant size.

The SIR model with demography extends the SIR model by also having births (or immigration) of susceptibles, at a fixed rate μN, and deaths (or emigration) from each state at rates μS, μI and μR respectively, where μ is the population turnover rate parameter. We note that this means the actual population size, $S+I+R$, is no longer fixed, but that the birth rate is held constant (i.e., N, the population size parameter, is constant). A consequence of the latter, along with the fact that the number of recovered individuals, R, has no direct bearing on the other states, and that our interest herein is on the number of infectious individuals, is that we may describe the state of the system by (S,I) (Kamenev and Meerson, 2008) with state space $\{(S,I) : 0 \le S,I\}$. The Markovian SIR model with demography we consider herein is detailed in Table 1 and Fig. 1.

2.2. Asymptotic approximations: the density process

We now state two limiting results of the SIR model with demography, in the limit as N becomes large (Kurtz, 1970, 1971; Pollett, 1990). These approximations assist us in defining the *probability of epidemic fade-out*, and are furthermore made use of in the two existing methods for approximating the probability of epidemic fade-out, discussed in Section 2.3, and in our own methods to be introduced in Section 3.

Let $Y_N(t)$ be a process following the model defined in Section 2.1, with each value being an (S,I) pair, and with initial value (S_0, I_0). The associated *density process* is $X_N(t) = Y_N(t)/N$, with each possible value x being an (s,i) pair, where $s = S/N$ and $i = I/N$; and the initial value is $x_0 = (s_0, i_0) = (S_0/N, I_0/N)$. The density process is important because it allows us to analyse the limiting behaviour as $N \to \infty$.

Let $f(x,l)$ be the transition rate of the density process from state (x) to state $(x+l/N)$, where l can take on the possible 1-step transition values in Table 1: $(-1,1)$, $(1,0)$, $(-1,0)$ and $(0,-1)$, respectively. Also define for the density process:

$$F(x) = \sum_l l f(x,l) = \big(-\beta si + \mu(1-s), \quad \beta si - (\gamma+\mu)i\big); \tag{1}$$

$B(x)$, a matrix whose (j,k)th element is given by $b_{j,k} = \frac{\partial f_j}{\partial x_k}$,

$$\Rightarrow B(x) = \begin{pmatrix} -\beta i - \mu & -\beta s \\ \beta i & \beta s - (\gamma+\mu) \end{pmatrix}; \tag{2}$$

Table 1
Events, transitions and their rates for the Markovian SIR model with demography.

Description	Transition	Rate
Infection	$(S,I) \to (S-1, I+1)$	$\beta SI/N$
Birth of susceptible	$(S,I) \to (S+1, I)$	μN
Removal of susceptible	$(S,I) \to (S-1, I)$	μS
Removal of infectious	$(S,I) \to (S, I-1)$	$(\gamma+\mu)I$

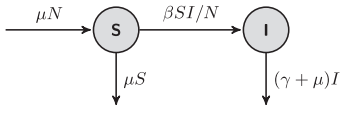


Fig. 1. Diagram of the Markovian SIR model with demography. Note, the "R" state is redundant and has been removed.

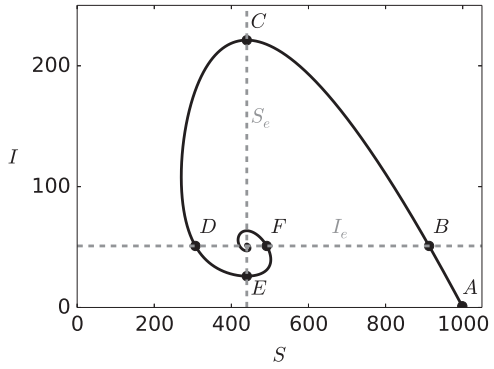


Fig. 2. A deterministic trace with $N=1000, \beta=2.5, \gamma=1, \mu=0.1$, and initial point (999, 1), with endemic values S_e and I_e (dashed lines). The outbreak starts at A, goes through points B, C, D and E through to F, and converges on (S_e, I_e). The *first trough* is between points D and F.

and $G(x)$, a matrix whose (j,k)th element is given by $g_{j,k}=\sum_l l_j\, l_k f(x,l)$,

$$\Rightarrow G(x)=\begin{pmatrix}\beta si+\mu(1+s) & -\beta si\\ -\beta si & \beta si+(\gamma+\mu)i\end{pmatrix}. \tag{3}$$

Then by Theorem 3.1 of Kurtz (1970) and Theorem 3.2 of Pollett (1990), we have *In the limit as* $N\to\infty$, $X_N(t)$ *weakly converges to a process which at time* t *is Gaussian with mean* $X(t)$ *and covariance* $\Sigma(t)/N$; *where* $X(t)$ *and* $\Sigma(t)$ *are the solutions to*

$$\frac{dX(t)}{dt}=F(X(t)),\quad X(0)=(s_0,i_0); \tag{4}$$

$$\frac{d\Sigma(t)}{dt}=B(X(t))\Sigma(t)+\Sigma(t)B(X(t))^T+G(X(t)),\quad \Sigma(0)=\mathbf{0}. \tag{5}$$

The solution to (4) when multiplied by N, $NX(t)$, is also known as the *deterministic approximation* to $Y_N(t)$. A typical solution of this approximation is shown in Fig. 2. The basic reproduction number corresponding to this approximation is

$$R_0=\beta/(\gamma+\mu). \tag{6}$$

Assuming $R_0>1$, the process begins at a point A, with a large number of susceptible individuals and a small number of infectious individuals (typically $A=(S,I)=(N-1,1)$). The outbreak rises through point B to point C, falls through point D to point E, and then rises again to point F. The cycle then repeats, on a smaller scale, as it spirals in towards the endemic value (S_e, I_e). Solving (1) for $F(x)=(0,0)$ and scaling by N give

$$(S_e,I_e)=N\left(\frac{(\gamma+\mu)}{\beta},\frac{\mu(\beta-\gamma-\mu)}{\beta(\gamma+\mu)}\right). \tag{7}$$

The maximum C and the minimum E of the deterministic infection curve occur when $S=S_e$, and we have chosen the points B, D and F to be at $I=I_e$.

2.3. Previous methods

The two main previous papers on this topic are by van Herwaarden (1997) and Meerson and Sasorov (2009). Both of these methods share in common the first part of the analysis: they assume that the outbreak follows the trajectory of the deterministic approximation until I falls below I_e (i.e., until point D in Fig. 2), and then set up a two boundary hitting probability problem. There is the natural lower absorbing boundary at $I=0$, and the probability of hitting this boundary before the upper boundary is used to approximate the probability of epidemic fade-out, here known as p_0. The choice of the upper absorbing boundary and method of solution of the two boundary hitting probability problem distinguishes the two methods.

van Herwaarden (1997)chooses the upper absorbing boundary to be $I=I_e$. He then approximates p_0 as the probability of hitting the $I=0$ boundary before the upper boundary. To do this, van Herwaarden assumes that S and I are continuous, then simplifies analysis by using the Fokker–Planck equation, effectively assuming the diffusion approximation as presented in Section 2.2. He uses boundary layer analysis to obtain an approximation to the probability of absorption at the lower boundary.

This results in the following approximation of p_0 (where W_0 is the principal branch of the Lambert W function, and Γ is the gamma function), assuming $1/\sqrt{N}\ll\mu\ll 1$:

$$\begin{aligned}
&x_{1A}=(-\gamma/\beta)W_0\left((-\beta/\gamma)\exp(-\beta/\gamma)\right),\\
&C_3=-\ln\left(\frac{-\beta x_{1A}}{\beta x_{1A}-\gamma}\right)\\
&\qquad-\int_{x_{1A}}^{1}\left(\frac{x_{1A}}{1-x_{1A}}\frac{\gamma(s-s\ln(s)-1)}{\beta s^2(1-s+(\gamma/\beta)\ln(s))}+\frac{1}{s-x_{1A}}\right)\mathrm{d}s,\\
&K=\frac{1}{\mu}\exp\left(\frac{\beta x_{1A}+(\beta-\gamma)\ln(1-x_{1A})}{\mu}+C_3\right),\\
&p_0=\exp\left(\frac{-KN\mu^2(\beta/\mu)^{(\beta-\gamma-\mu)/\mu}\exp(-\beta/\mu)}{(\gamma+\mu)\Gamma((\beta-\gamma-\mu)/\mu)}\right).
\end{aligned} \tag{8}$$

Meerson and Sasorov (2009) instead use a slightly different upper absorbing boundary, namely a diagonal line from (S_e, I_e) to $(N,0)$, and employ the WKB (Wentzel–Kramers–Brillouin) expansion method (Bender and Orszag, 1978) in place of the Fokker–Planck equation to solve the two boundary problem. This results in the following approximation of p_0:

$$\begin{aligned}
&K=\beta/\mu,\\
&\delta=1-(\gamma+\mu)/\beta,\\
&x_m=(-(\gamma+\mu)/\beta)W_0\left((-\beta/(\gamma+\mu))\exp(-\beta/(\gamma+\mu))\right)-1,\\
&Q_1=\int_0^{x_m}\left(\frac{s(s+\delta)}{(1+s)^2(s-(1-\delta)\ln(1+s))}-\frac{x_m}{(1+x_m)(s-x_m)}\right)ds,\\
&y_m=\frac{(\delta+x_m)x_m}{1+x_m}\left(\frac{-x_m}{\delta}\right)^{K\delta}\exp\left(K(x_m+\delta)-(1+x_m^{-1})Q_1\right),\\
&C=\frac{y_m\delta}{2\pi(1-\delta)},\\
&S_0=C\sqrt{\frac{2\pi}{K\delta}},\\
&p_0=\exp(-NS_0).
\end{aligned} \tag{9}$$

The analysis assumes $NS_0\gg 1$, and hence p_0 close to 0; however it turns out to be quite accurate in nearly all cases.

We assess the accuracy of these approximations in Section 4.1. However, we note that a nice property of these methods is that they give explicit mathematical expressions for p_0, with negligible computing time.

3. New algorithms

We now consider new algorithms for approximating the probability of epidemic fade-out. These methods are based upon reducing the dependency upon asymptotic approximations, yet still seeking to retain computational efficiency.

Similar to the existing methods, we decompose the problem into two parts. We first consider the state of the process upon its first entrance to the first trough. In place of the deterministic approximation used in the earlier work, we calculate an approximate distribution of the process based upon the diffusion approximation as presented in Section 2.2; further details are presented below in Section 3.2.

In the second part of our methods, we calculate the probability of reaching $I=0$ before exiting the first trough. In this region, where I is relatively small and stochastic effects are important, we use discrete-state stochastic models. We present two alternative ways to do this: an exact computation in Section 3.3.1, and an efficient approximation in Section 3.3.2. This theoretically should further improve upon previous methods, which used asymptotic approximations (diffusion or WKB) for this calculation. In our discrete-state, stochastic representation of the system, the definition of the first trough can be unclear; for this reason, we commence by providing a precise definition of epidemic fade-out, which we adopt in our methods.

3.1. Definition of epidemic fade-out

Since we are dealing with a discrete system, we round up the endemic fixed point to the next highest integer pair, i.e. $S_d=\lceil S_e\rceil$ and $I_d=\lceil I_e\rceil$. Let $T=(S,I):S<S_d, I=I_d$, i.e. the set of states on the dotted (green) line in Fig. 3. We define the entrance to the first trough as the point when the system first enters T.

It is now tempting to define the end of the first trough as the point when the system next reaches $I>I_d$. However, due to stochastic effects, it is possible for the system to immediately jump up to $I>I_d$, but this certainly does not mean the end of the first trough. Therefore we need to set the upper boundary to a value greater than I_d in the $S<S_d$ region. So long as it is sufficiently far from I_d to avoid small fluctuations, this value is not critical. Hence $2I_d$ was chosen because it means that T is an equal distance from each absorbing I boundary.

So we define the end of the first trough to be when the system reaches either $I=2I_d$, or both $S\geq S_d$ and $I\geq I_d$; i.e., the dashed (red) lines in Fig. 3.

We then define p_0, the probability of epidemic fade-out, as the probability that the process is absorbed at $I=0$ before it exits the first trough (that is, before it reaches the dashed red line), given that the process reaches the start of the first trough.

3.2. Part I: The distribution upon first entrance

In the first part of our method we approximate the distribution of the process upon its first entrance to T, i.e., the distribution of the process upon first reaching the green dotted line in Fig. 3. Results of Ethier and Kurtz (1986) provide this distribution, corresponding to the diffusion approximation presented in Section 2.2.

Let τ be the time at which $X(t)$ first enters T, i.e. $\tau=\min\{t\geq 0: X(t)\in T\}$.

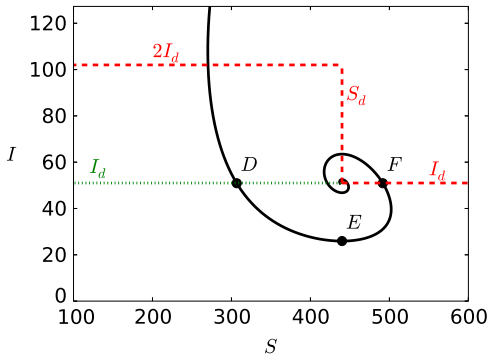


Fig. 3. Deterministic trace with the same parameters, and same points D, E and F, as Fig. 2. States in T, which denotes the start of the first trough, are shown by the dotted (green) line. The end of the first trough is shown by the dashed (red) line.

We may then approximate the "hitting distribution" of $X(t)$ when it first enters T, as now described.

If we use the subscript j to denote the j^{th} element of a vector, and subscript j,k to denote the row j, column k element of a matrix; then by applying Theorem 11.4.1 of Ethier and Kurtz (1986) we have: *In the limit as $N\to\infty$, the distribution of the density process $X_N(t)$ when it first enters T is Gaussian, with mean $X(\tau)_1$ and variance*

$$H=\left(\Sigma(\tau)_{1,1}+\left(\frac{F(X(\tau))_1}{F(X(\tau))_2}\right)^2\Sigma(\tau)_{2,2}-2\left(\frac{F(X(\tau))_1}{F(X(\tau))_2}\right)\Sigma(\tau)_{1,2}\right)/N; \quad (10)$$

where $F(x)$, $X(t)$ and $\Sigma(t)$ are as defined in Eqs. (1), (4) and (5). We start the diffusion at point B in Fig. 2 (that is, $\Sigma(t)=\mathbf{0}$ at point B) because we condition on a major outbreak occurring.

In our methods, we approximate the hitting distribution of S in T by discretising a Gaussian distribution with mean $NX(\tau)_1$ and variance N^2H, and renormalising such that $S\geq 0$ and $S<S_d$ (because $S\geq S_d$ corresponds to sample paths which never meet the criteria of entering the first trough). We call this discrete, renormalised distribution Δ, and it is the initial distribution for the calculations in Sections 3.3.1 and 3.3.2.

3.3. Part II: Modelling the behaviour within the first trough

For the second part of the computation we define the first trough Markov chain, with states arranged as in Fig. 4. Each state is represented by an (S,I) pair. Column S has states $(S,0)$ to $(S,2I_d)$ for $S\leq S_d$, and has states $(S,0)$ to (S,I_d) for $S>S_d$. There are two absorbing boundaries, representing the two possible outcomes: the lower absorbing boundary corresponds to epidemic fade-out occurring, and consists of the states $(S,0)$ for all S; the upper absorbing boundary corresponds to epidemic fade-out not occurring, and consists of the states $(S,2I_d)$ for $S<S_d$, (S,I_d) for $S>S_d$, and (S,n), for all $n\in[I_d,2I_d]$, for $S=S_d$.

3.3.1. Exact model

We can evaluate p_0 exactly (for a given starting distribution, Δ). One way to solve this, using standard techniques (Norris, 1998), would be to simultaneously solve equations for all the approximately $(N+S_d)\times I_d$ points in the first trough. However we may take advantage of the fact that almost invariably $2I_d\ll N$, and solve it more efficiently by evaluating a column at a time, as now described.

For each column S, create a stochastic transition matrix, A_S, of the first exit from column S to its neighbouring columns. We partition A_S into two matrices F_S and B_S – such that $A_S=[F_S B_S]$ –

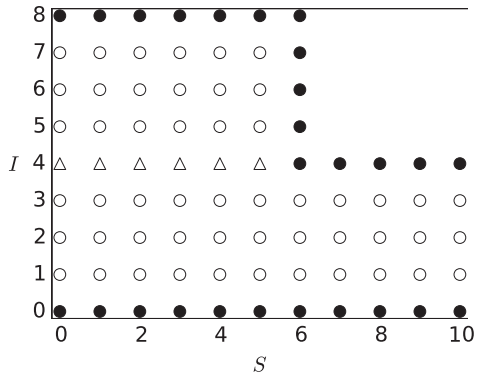


Fig. 4. A two-dimensional representation of the states in the first trough Markov chain, with $S_d=6$ and $I_d=4$. The states in the absorbing boundaries are shown as solid circles. States in T (the start of the first trough) are denoted by triangles, and contain Δ, the distribution of the process when it enters the first trough. States continue infinitely to the right.

which represent the first exit into the next and previous column, respectively. Provided state (S,m) is not in the upper absorbing boundary: then the (m,n)th element of F_S is the probability that the first exit from column S is into state $(S+1,n)$, given that the process starts in state (S,m); and the (m,n)th element of B_S is the probability that the first exit from column S is into state $(S-1,n)$, given that the process starts in state (S,m). These probabilities are calculated exactly for the Markovian SIR model with demography, which has the transition rates illustrated in Fig. 5(a).

The upper absorbing boundary is an artificial absorbing boundary, so we need to treat it in a specific manner. We define F_S and B_S such that any probability mass in the upper absorbing boundary of column S moves to the upper absorbing boundary of column $S+1$. More specifically, if the state (S,m) is in the upper absorbing boundary, then row m of $[F_S B_S]$ is all zeros, except the $(m,n)^{\text{th}}$ element of F_S which is equal to 1, where $n=2I_d$ if $S<S_d$, or $n=I_d$ if $S\geq S_d$.

We define P_S as the matrix of *first entry* into column $S+1$ from column S. So the $(m,n)^{\text{th}}$ element of P_S is the probability that the first entry to column $S+1$ is into state $(S+1,n)$, given that the process starts in state (S,m).

With these definitions, we establish the recursive relation:

$$P_S=\begin{cases} F_0 & \text{if } S=0 \\ (\boldsymbol{I}-B_S P_{S-1})^{-1}F_S & \text{if } S>0, \end{cases} \tag{11}$$

where $\boldsymbol{I}$ (here only) is the identity matrix. The F_S and B_S matrices are straightforward to calculate, so we can determine each P_S matrix.

Now define the vector D_S to be the distribution of Δ in column S. The only non-zero element of D_S is element I_d, and then only if $S<S_d$. Finally, we define E_S to be the distribution of all probability mass which first entered the first trough at column S or less, on the first trough Markov chain's first entry to column S. By definition, all probability mass enters the first trough at $S<S_d$. So for $S\geq S_d$, the definition of E_S simplifies to: the probability distribution on the first trough Markov chain's first entry to column S.

We then use P_S and D_{S+1} to calculate E_{S+1} through the recursion:

$$\begin{aligned} &E_0=D_0, \\ &E_{S+1}=E_S P_S+D_{S+1} \quad \text{if } S\geq 0. \end{aligned} \tag{12}$$

As we increment S, eventually all but a vanishingly small amount of the probability mass is at one of the absorbing boundaries. That is, for any δ, there is an $S\geq S_d$ such that elements 0 and I_d of E_S sum to greater than $1-\delta$; when this occurs, p_0 is taken to be element 0 of E_S.

For a given Δ, this method gives an *exact* result (to the accuracy of the δ chosen). But (as we shall see in Section 4.1) it does not scale well to very large population sizes. This is because it requires the calculation and storage of four matrices (F_S, B_S, P_{S-1} and P_S) with a maximum size of $(2I_d+1)\times(2I_d+1)$. We now proceed to consider an approximate model which reduces the computational overheads.

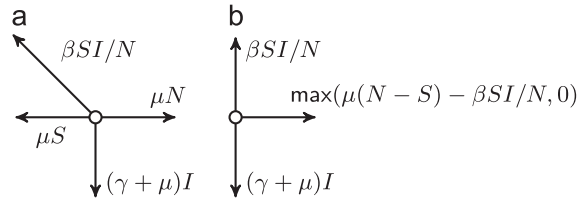


Fig. 5. Transition rates from state (S,I) (except when state (S,I) is in the upper absorbing boundary. (a) shows the exact model (Section 3.3.1). (b) shows the approximate model (Section 3.3.2), in which every infection event (rate $\beta SI/N$) or removal of susceptible event (rate μS) is "paired" with a birth event (rate μN).

3.3.2. Approximate model

Considering the method in Section 3.3.1, we can substantially reduce both evaluations and storage by making the following observation: whenever the system goes back from column S to column $S-1$, it returns to column S after an unknown number of intermediate events, followed by a "birth" event (because the birth events, at rate μN, are the only events which increase the number of susceptibles). So let us – as an approximation – assume that this "unknown number of intermediate events" is in fact *no events*. In other words, every "death of a susceptible" event (at rate μS) is paired with a birth event (at rate μN); and every infection event (at rate $\beta SI/N$) is also paired with a birth event. So, with this assumption, we calculate the jump chain transition probabilities using the one step transition rates shown in Fig. 5(b) rather than Fig. 5(a).

If we were to follow the analysis of Section 3.3.1, that would mean $B_S=0$ and so Eq. (11) reduces to $P_S=F_S$. However it is possible to avoid generating F_S (or any other large matrices) altogether.

Given E_S and the transition probabilities within column S, we can calculate the expected number of visits to each state before exiting the column. Since each state (S,I) only communicates with the adjacent states $(S,I-1)$ and $(S,I+1)$, this is a tri-diagonal series of simultaneous equations, which can be solved using an efficient technique such as the Thomas algorithm (Hoffman, 2001). These expected numbers of visits multiplied by the transition probabilities to the right (determined from the transition rates in Fig. 5 (b)) give E_{S+1}. In other words, with reference to Eq. (12), we calculate the vector $E_S P_S$ directly without calculating P_S.

The elements of E_S corresponding to the absorbing boundaries ($I=0$, and $I=2I_d$ or I_d) accumulate probability mass as S increments. As in Section 3.3.1, for sufficiently large $S\geq S_d$, all but a vanishingly small amount of probability mass is absorbed, allowing us to efficiently approximate p_0.

Given the simpler computation and low storage requirement, it is no surprise that this is much faster than the exact method in Section 3.3.1. But, as we shall show in the next section, this method is also very accurate.

4. Results

4.1. Accuracy and efficiency

In this section we compare the accuracy of all methods – the previous work of van Herwaarden (1997) and Meerson and Sasorov (2009) as presented in Section 2.3, and our new algorithms as detailed in Section 3.

As references, we also add an exact computation, and a Monte Carlo simulation. The exact computation uses a truncated state space, with an extra absorbing boundary at $S+I=(1.1)N$. The amount of probability mass absorbed at that boundary is extremely low (never greater than 10^{-5}) and does not affect the results significantly. For both the exact computation and Monte Carlo simulation, as well as our methods, we calculate p_0 using the definition in Section 3.1. For previous works, we use the expressions in Section 2.3.

Since we are interested in evaluating a probability which is only state (and not time) dependent, we may scale time and hence fix $\gamma=1$, so time is in units of the average infectious period of an individual.

We chose seven values of the population size parameter, N: 1000, 3000, 10,000, 30,000, 100,000, 300,000 and 1,000,000; and

six effective transmission rate parameter values, β:1.1,1.2,1.5,2, 4 and 8. For each of these 42 pairs of values, we chose three values of population turnover rate parameter μ, to give final p_0 values of approximately 0.1, 0.5 and 0.9 respectively. In a few cases (notably for low N and low β), an appropriate μ value could not be found. The μ values used are given in Table A1 in Appendix A.

For $N \leq 3000$ all methods were compared against an exact computation. For larger N they were all compared against Monte Carlo simulations. The worst case and average errors for each N value are shown in Fig. 6. Error bars are due to the uncertainty in the result from the Monte Carlo simulations, which require a very long run time for large N.

We see that both our methods are noticeably more accurate than the previously published results. We also see the accuracies of our two methods are comparable. This suggests that most of the error is in the diffusion approximation of Δ (Section 3.2), rather than the Markov chain approximation of Section 3.3.2.

We also consider the efficiency of our methods, and compare them to exact computations and Monte Carlo simulations. To give comparable (though generally lower) accuracy, the Monte Carlo simulations were run long enough to give a standard deviation of $\sigma = 0.005 = 0.5\%$ in their approximations of p_0. All tests were run on a 2014 iMAC (Intel i5 core, 2.7 GHz, 8 GB RAM, Mac OS X) running Cython (Behnel et al., 2011). Computation time for different methods is shown in Fig. 7. The methods of van Herwaarden and Meerson and Sasorov are not shown, because they take negligible time (in the milliseconds), and their times are independent of N. Therefore these methods remain the best for getting an approximate answer quickly.

The time for the exact computation is approximately proportional to N^3, and quickly becomes impractical. Our first method, based upon a diffusion approximation to the first entrance to the first trough, combined with an exact computation (Section 3.3.1) also has a computation time approximately proportional to N^3, and becomes impractical as N approaches 10^5.

Our second method, based upon a diffusion approximation to the first entrance to the first trough, combined with an approximate model (Section 3.3.2) has a time which is approximately proportional to N^2 and so is practical up to at least $N = 10^7$. The small memory overhead means even larger sizes may be computed if a long run time is not a concern.

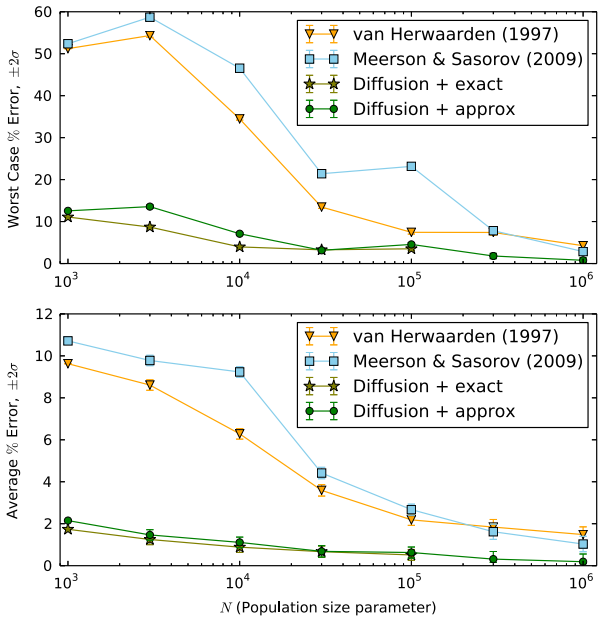


Fig. 6. Plot of worst case and average error versus population size N, with $\pm 2\sigma$ error bars.

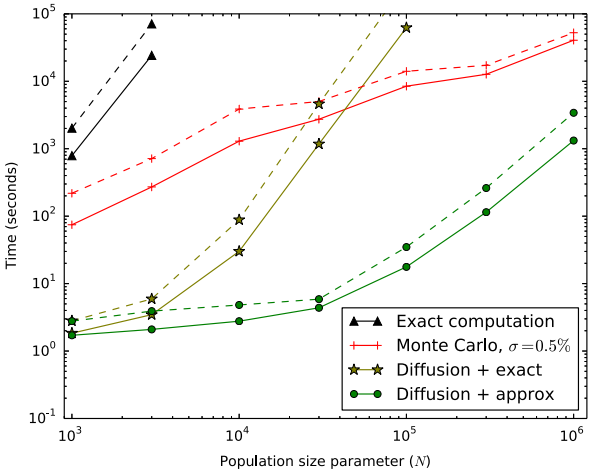


Fig. 7. Plot of computation time versus population size N. Mean computation times are shown with solid lines. Slowest computation times are shown with dashed lines.

The time for a set of Monte Carlo simulations is proportional to a little less than N^2, though our method is projected to be faster and more accurate up to at least $N = 10^7$.

So for a very wide range of N (from a few thousand, to the millions), the diffusion plus approximate model algorithm appears to be the most accurate of practical methods.

4.2. Analysis of the results

We used the method of Section 3.3.2 to run a larger set of tests, to explore how the probability of epidemic fade-out changes as a function of model parameters.

For $N = 1000, 10,000, 100,000$ and 1,000,000, we tested: $\gamma = 1$; β values from 1.1 to 5, stepping in increments of 0.1; and μ values from 0.010 up to 0.089, 0.049, 0.029 and 0.019 for the respective values of N, stepping in increments of 0.001. Contour plots of the p_0 values are shown in Fig. 8.

Fig. 8 shows the interesting result that p_0 is generally non-monotonic in β. Naively, one might expect a higher infection rate β to cause the infection to be more persistent, and so give a lower p_0. What we instead see, in most cases, is a local maximum near $\beta = 2$. Since $\mu \ll 1$ and $\gamma = 1$, it follows that $R_0 = \beta/(\gamma+\mu) \approx \beta$ and so the local maximum is also near $R_0 = 2$.

The main reason for the non-monotonicity, and the peak near $R_0 = 2$, can be seen by considering the traces for $R_0 = 5$, 2 and 1.3 in Fig. 9. Note these are the solutions to Eq. (4) (scaled by N), and hence the deterministic approximations to the epidemic. We define I_m to be the minimum I value in the first trough of the deterministic approximation.

We may rearrange Eqs. (1) and (7) to give $dI/dt = (\beta I/N)(S - S_e)$ and hence $d(\ln(I))/dt = (\beta/N)(S - S_e)$. This means that for a given deterministic curve, the rate of change of $\ln(I)$ is proportional to $S - S_e$. It can also be shown that the minimum S occurs at $I > I_e$.

If we compare the $R_0 = 2$ curve to the $R_0 = 1.3$ curve, we see that the $R_0 = 2$ curve starts further from its endemic point; that is, it has a higher initial $S - S_e$ value. This means $S - S_e$ is higher in the early stages of the outbreak, which causes I to rise more steeply

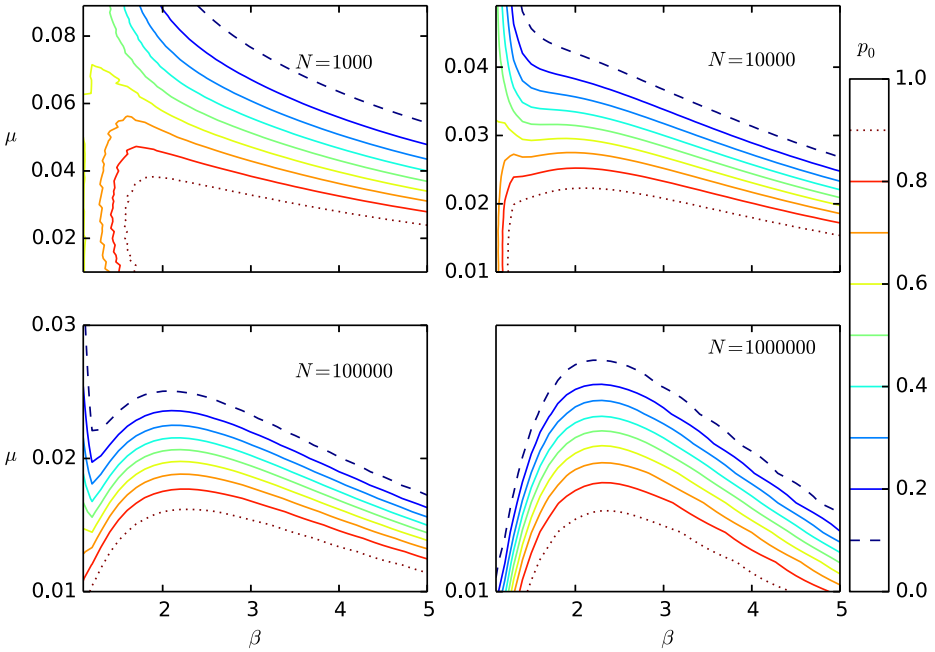


Fig. 8. Constant p_0 contours for various β, μ and N values, with $\gamma = 1$ and $(S_0, I_0) = (N-1, 1)$. The $p_0 = 0.1$ contour is dashed, the $p_0 = 0.9$ contour is dotted, and the contours for $p_0 = 0.2$–0.8 in steps of 0.1 are solid. For most combinations of N and μ, p_0 peaks near $\beta = 2$.

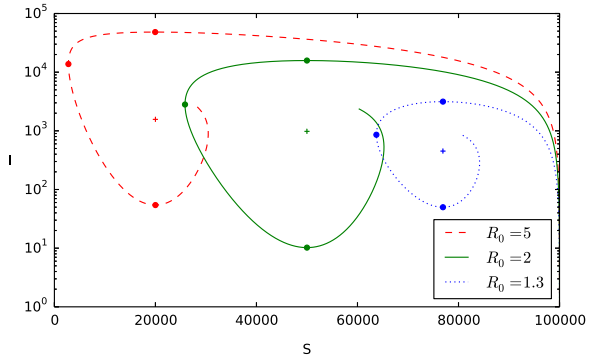


Fig. 9. Comparison of deterministic traces, on a logarithmic I scale, for $N = 100,000$, $\gamma = 1$, $\mu = 0.02$ and $(S_0, I_0) = (N-1, 1)$. The R_0 values of 5, 2 and 1.3 correspond to S_e values of approximately $N/5$, $N/2$ and $4N/5$ respectively. The plus signs show the endemic points (S_e, I_e) to which the curves converge. The dots mark the maximum I, minimum S and first trough minimum $I(I_m)$.

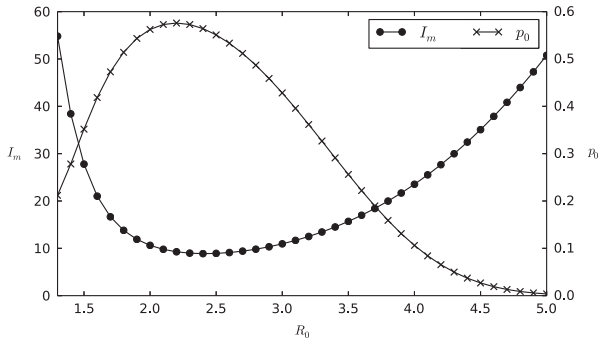


Fig. 10. I_m and p_0 plotted against R_0, for $N = 100,000$, $\gamma = 1$, $\mu = 0.02$, $(S_0, I_0) = (N-1, 1)$, and $\beta = R_0(\gamma + \mu)$. In this example, the minimum I_m is at $R_0 \approx 2.4$, while the maximum p_0 is at $R_0 \approx 2.2$.

and for longer, so the maximum $I - I_e$ value is higher. This in turn gives S more time to fall, so the curve reaches a lower minimum $S - S_e$ value. Finally, this gives more time for I to fall, so the $R_0 = 2$ curve falls to a considerably lower I_m value then the $R_0 = 1.3$ curve. Biologically, the $R_0 = 2$ case indicates the infection "burning out" – it uses up so much resources (indicated by S falling low) that it is slow to re-establish itself, so it falls to a low I_m, giving it a higher probability of epidemic fade-out.

If we compare the $R_0 = 5$ curve to the $R_0 = 2$ curve, the $R_0 = 5$ curve has an even higher initial $S - S_e$ value, and so rises to a higher maximum $I - I_e$. But then, as S falls, it is limited by the condition that $S \geq 0$. So it does not fall to as low a minimum $S - S_e$ value as the $R_0 = 2$ curve does. This in turn means that its I_m is not as low as for $R_0 = 2$. It also only has to reach a comparatively low S before $S > S_e$, and the deterministic curve begins to rise again. Biologically, $R_0 = 5$ corresponds to a case where the infection rate is so high that the infection can re-establish itself from comparatively low resources.

So $S_e \approx N/2$ corresponds to a "sweet spot" where the curve can swing from a high $S - S_e$ to a low $S - S_e$, giving the most time for the curve to fall to a low I_m. And it follows from (6) and (7) that $R_0 = 2$ corresponds to $S_e = N/2$, and so this sweet spot occurs near $R_0 = 2$. This is illustrated in Fig. 10, which plots I_m versus R_0 for the same parameters as used in Fig. 9, with the lowest I_m occurring at $R_0 \approx 2.4$.

When we consider stochastic effects, a lower I_m corresponds to a greater probability of absorption at $I = 0$, and hence a higher p_0. However the probability of absorption also depends on the time the deterministic process spends near I_m. A longer time near I_m

corresponds to a longer time near the $I=0$ absorbing boundary. This gives the process more opportunity to be absorbed due to stochastic effects, and so should lead to a higher p_0.

It is shown in Appendix B that in the first trough of the deterministic process, the time for which $I < I_m + \epsilon$, for sufficiently small ϵ, is monotonically decreasing in βI_m. This means it is also monotonically decreasing in $R_0 I_m$. We must handle this result with some care because I_m is itself dependent on R_0. But it means that, in the region where the I_m versus R_0 curve is relatively flat, a decrease in R_0 gives an increase in p_0. So we would expect the maximum p_0 to occur at an R_0 slightly lower than the R_0 which gives rise to the minimum I_m. This is also illustrated in Fig. 10. For this particular case the minimum I_m occurs at $R_0 \approx 2.4$, but the lowest p_0 occurs at $R_0 \approx 2.2$.

5. Conclusion

We have presented a two stage method for calculating an accurate approximation for the probability of epidemic fade-out. Using an approximate model on the second stage gives an algorithm which is both fast and accurate. It is more accurate than the previously published formulae, and scales much better than exact computation methods. This technique can also be used in other SIR-type models with replenishment of susceptibles (for instance, those with a fixed population size).

A possible justification for why the approximate model of Section 3.3.2 retains such a high level of accuracy is that in the first trough (when I is low), the birth events (μN) are almost always at a higher rate – and often a much higher rate – than infection events ($\beta SI/N$). Therefore there is a very small penalty (in terms of accuracy) for pairing every infection event with a birth event.

Further, comparing Fig. 5(b) to (a), we see that the approximation makes no change to the one step behaviour in the I (vertical) dimension. In the S (horizontal) dimension, the behaviour is simplified, but the average drift $(\mu(N-S)-\beta SI/N)$ is modelled correctly (except when $\mu(N-S)-\beta SI/N<0$, but in those cases the Markov chain is near point F in Fig. 3, so the computation is nearly complete). So the S dimension is modelled accurately in the first moment but not the second moment. It appears that this only introduces a small error because the I dimension is much more critical than the S dimension.

Using this fast and accurate method, we have found that the probability of epidemic fade-out often peaks when the basic reproduction number, R_0, is approximately 2 (restricting consideration to cases where a major outbreak is possible, i.e., $R_0 > 1$). This is because $R_0 \approx 2$ is high enough to use up a large proportion of resources, but not so high that the infection can easily recover from having few resources. The reason this occurs near $R_0 = 2$ appears to be due to the endemic point being near $S = N/2$.

A potential public health application is that there may be instances where action against an infection should be limited, to maximise the chance of infection being eliminated before it becomes endemic. We note there is some similarity here to the observations of Rozhnova et al. (2013), that decreasing R_0 by vaccination may sometimes lead to higher persistence, though their study was with respect to an already endemic infection, with seasonality.

The question of whether a peak near $R_0 = 2$ extends to other measures or models, is a topic for future research. Another avenue for future research is to determine methods which allow calculation of the probability of epidemic fade-out for models with seasonal forcing (i.e., a time-dependent effective transmission rate parameter) (Keeling et al., 2001). This in turn could aid understanding of the Critical Community Size for diseases such as measles in the pre-vaccine era (Conlan et al., 2010).

6. Acknowledgements

The authors thank Nic Rebuli for helpful discussions on the diffusion approximation (Section 3.2). We also thank Guy Latouche for a suggestion which led us to the efficient and exact solution in Section 3.3.1. This work is supported by an APA Scholarship (PB), an Australian Research Council Future Fellowship (JVR; FT130100254), and the NHMRC (JVR; CRE PRISM^2).

Appendix A

See Table A1.

Table A1
μ values used in Section 4.1.

N	β	μ for $p_0 \approx 0.9$	μ for $p_0 \approx 0.5$	μ for $p_0 \approx$ 0.1
1000	1.1	–	0.43	–
	1.2	0.053	–	–
	1.5	0.034	0.084	–
	2.0	0.033	0.060	0.112
	4.0	0.025	0.041	0.064
	8.0	0.015	0.025	0.039
3000	1.1	–	0.064	–
	1.2	0.026	0.050	–
	1.5	0.025	0.046	0.091
	2.0	0.026	0.041	0.062
	4.0	0.021	0.031	0.042
	8.0	0.012	0.018	0.025
10,000	1.1	0.020	0.051	–
	1.2	0.017	0.038	0.095
	1.5	0.019	0.030	0.046
	2.0	0.021	0.031	0.041
	4.0	0.017	0.024	0.031
	8.0	0.010	0.014	0.018
30,000	1.1	0.012	0.030	–
	1.2	0.012	0.022	0.041
	1.5	0.016	0.023	0.031
	2.0	0.018	0.025	0.031
	4.0	0.015	0.020	0.024
	8.0	0.008	0.011	0.014
100,000	1.1	0.008	0.016	0.033
	1.2	0.009	0.015	0.022
	1.5	0.013	0.018	0.023
	2.0	0.015	0.020	0.025
	4.0	0.013	0.017	0.020
	8.0	0.007	0.010	0.012
300,000	1.1	0.006	0.010	0.017
	1.2	0.008	0.012	0.016
	1.5	0.011	0.015	0.019
	2.0	0.014	0.017	0.021
	4.0	0.011	0.015	0.017
	8.0	0.006	0.008	0.010
1,000,000	1.1	0.005	0.007	0.010
	1.2	0.006	0.009	0.012
	1.5	0.010	0.013	0.015
	2.0	0.012	0.015	0.018
	4.0	0.010	0.013	0.015
	8.0	0.006	0.007	0.008

Appendix B

theorem. *In the first trough of the deterministic process, the time θ for which $I < I_m + \epsilon$, for sufficiently small ϵ, is monotonically decreasing in βI_m.*

Proof. Consider the deterministic plot of I versus S, as in Figs. 2 and 3. At the first trough minimum (point E), $dI/dS = 0$. For sufficiently small ϵ, we can therefore treat d^2I/dS^2 as constant, and in the region where $I < I_m + \epsilon$, I is parabolic when plotted against S. So the distance in the S dimension, for which $I < I_m + \epsilon$, is monotonically decreasing in the parabola curvature d^2I/dS^2.

The rate at which the deterministic process moves in the S direction is dS/dt, so θ is inversely proportional to dS/dt. This means that θ is monotonically decreasing in $(d^2I/dS^2)(dS/dt)$.

Substituting $I = Ni$ and $S = Ns$ into (1) gives

$$\frac{dS}{dt} = \mu N - \mu S - \beta SI/N, \tag{B.1}$$

$$\frac{dI}{dS} = \frac{dI/dt}{dS/dt} = \frac{\beta SI/N - (\gamma+\mu)I}{\mu N - \mu S - \beta SI/N} \tag{B.2}$$

$$\Rightarrow \frac{d^2I}{dS^2} = \frac{(\beta I/N)(\mu N - \mu S - \beta SI/N) - [\beta SI/N - (\gamma+\mu)I](-\mu - \beta I/N)}{(\mu N - \mu S - \beta SI/N)^2}$$

$$= \frac{I\mu(\beta - \gamma - \mu) - \beta I^2(\gamma+\mu)/N}{(\mu N - \mu S - \beta SI/N)^2}. \tag{B.3}$$

At the first trough minimum of the deterministic curve we have defined $I = I_m$. Also $dI/dS = 0$, so it follows from (B.2) that $S = N(\gamma+\mu)/\beta$, and $N - S = N(\beta - \gamma - \mu)/\beta$. Substituting these into Eqs. (B.1) and (B.3) gives

$$\frac{dS}{dt} = \mu N(\beta - \gamma - \mu)/\beta - I_m(\gamma+\mu),$$

$$\frac{d^2I}{dS^2} = \frac{I_m\mu(\beta - \gamma - \mu) - \beta I_m^2(\gamma+\mu)/N}{[\mu N(\beta - \gamma - \mu)/\beta - I_m(\gamma+\mu)]^2};$$

$$\Rightarrow \left(\frac{d^2I}{dS^2}\right)\left(\frac{dS}{dt}\right) = \frac{\beta I_m}{N}. \tag{B.4}$$

Therefore θ is monotonically decreasing in βI_m.□

References

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, New York.

Andersson, H., Britton, T., 2000. Stochastic Epidemic Models and Their Statistical Analysis. Springer-Verlag, New York.

Ball, F.G., 1983. The threshold behaviour of epidemic models. J. Appl. Probab. 20 (2), 227–241.

Ball, F.G., Lyne, O.D., 2002. Optimal vaccination policies for stochastic epidemics among a population of households. Math. Biosci. 177, 333–354.

Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D.S., Smith., K., 2011. Cython: the best of both worlds. Comput. Sci. Eng. 13 (2), 31–39.

Bender, C., Orszag, S., 1978. Advanced Mathematical Method for Scientists and Engineers. McGraw-Hill, New York.

Black, A.J., House, T., Keeling, M.J., Ross, J.V., 2013. Epidemiological consequences of household-based antiviral prophylaxis for pandemic influenza. J. R. Soc. Interface 10, 20121019.

Britton, T., House, T., Lloyd, A.L., Mollison, D., Riley, S., Trapman, P., 2015. Five challenges for stochastic epidemic models involving global transmission. Epidemics 10, 54–57.

Conlan, A.J.K., Rohani, P., Lloyd, A.L., Keeling, M., Grenfell, B.T., 2010. Resolving the impact of waiting time distributions on the persistence of measles. J. R. Soc. Interface 7, 623–640.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. John Wiley & Sons, Chichester.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes. John Wiley & Sons, Hoboken.

Hoffman, J.D., 2001. Numerical Methods for Scientists and Engineers, 2nd edition. Marcel Dekker, New York.

Kamenev, A., Meerson, B., 2008. Extinction of an infectious disease: a large fluctuation in a nonequilibrium system. Phys. Rev. E 77, 061107.

Keeling, M.J., Rohani, P., Grenfell, B.T., 2001. Seasonally forced disease dynamics explored as switching between attractors. Phys. D: Nonlinear Phenom. 148 (3–4), 317–335.

Kurtz, T.G., 1970. Solutions of ordinary differential equations as limits of pure jump Markov processes. J. Appl. Probab. 7 (1), 49–58.

Kurtz, T.G., 1971. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. J. Appl. Probab. 8 (2), 344–356.

McCaw, J.M., McVernon, J., 2007. Prophylaxis or treatment? Optimal use of an antiviral stockpile during an influenza pandemic. Math. Biosci. 209, 336–360.

Meerson, B., Sasorov, P.V., 2009. WKB theory of epidemic fade-out in stochastic populations. Phys. Rev. E 80, 041130.

Nåsell, I., 1999. On the time to extinction in recurrent epidemics. J. R. Stat. Soc.: Ser. B 61, 309–330.

Nåsell, I., 2001. Extinction and quasi-stationarity in the Verhulst logistic model. J. Theor. Biol. 211, 11–27.

Norris, J.R., 1998. Markov Chains. Cambridge University Press, Cambridge, UK.

Pollett, P.K., 1990. On a model for interference between searching insect parasites. J. Aust. Math. Soc. Ser. B 32, 133–150.

Ross, J.V., Black, A.J., 2015. Contact tracing and antiviral prophylaxis in the early stages of a pandemic: the probability of a major outbreak. Math. Med. Biol. 32 (3), 331-343.

Rozhnova, G., Metcalf, C.J. E., Grenfell, B.T., 2013. Characterizing the dynamics of rubella relative to measles. J. R. Soc. Interface 10, 20130643.

van Herwaarden, O.A., 1997. Stochastic epidemics: the probability of extinction of an infectious disease at the end of a major outbreak. J. Math. Biol. 35, 793–813.

van Herwaarden, O.A., Grasman, J., 1995. Stochastic epidemics: major outbreaks and the duration of the endemic period. J. Math. Biol. 33, 581–601.

4 Paper 2

4.1 Introduction

Paper 2 is entitled "Intervention to maximise the probability of epidemic fade-out". It was published in *Mathematical Biosciences* in 2017 [7].

Paper 1 had made the observation that p_0 was non-monotonic in β, the transmission rate parameter. This suggested that there were two or more competing effects influencing epidemic fade-out.

We assumed the existence of a method of control which allows one to reduce β. Then this paper investigated how to find the optimal policy to apply that control, in order to maximise p_0. The analysis was performed for the SIR-with-demography model, under two different control scenarios.

We first used Markov decision theory, which provides optimal policies, but is impractical to use for all but the smallest systems. Then the main contribution of the paper is that we derived a simple formula which gives a close to optimal policy. This formula is simple, and independent of the chosen values of β. We also demonstrated that close to optimal results are obtained even if only an approximation of this policy can be enforced.

4.2 Statement of Authorship

Statement of Authorship

Title of Paper	Intervention to maximise the probability of epidemic fade-out
Publication Status	**Published**
Publication Details	P. G. Ballard, N. G. Bean, J. V. Ross, Intervention to maximise the probability of epidemic fade-out, Mathematical Biosciences, 293:1-10, November 2017, doi:10.1016/j.mbs.2017.08.003

Principal Author

Name of Principal Author (Candidate)	Peter Ballard		
Contribution to the Paper	Derived all the mathematics, wrote all the algorithms and code, ran all the code, generated all the diagrams, wrote most of the text		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20-3-2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Prof. Nigel Bean		
Contribution to the Paper	Project supervision, idea generation, suggestions and corrections to text.		
Overall percentage (%)	10%		
Signature		Date	28/03/2018

Name of Co-Author	Prof. Joshua Ross		
Contribution to the Paper	Project supervision, idea generation, suggestions and corrections to text.		
Overall percentage (%)	10%		
Signature		Date	28/03/18

4.3 Paper 2

The paper, as published in *Mathematical Biosciences* in 2017, is on the following pages.

Mathematical Biosciences 293 (2017) 1–10

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Intervention to maximise the probability of epidemic fade-out

P.G. Ballard*, N.G. Bean, J.V. Ross

The University of Adelaide, School of Mathematical Sciences and ARC Centre of Excellence for Mathematical and Statistical Frontiers, Adelaide SA 5005, Australia

ARTICLE INFO

Article history:
Received 19 October 2016
Revised 4 August 2017
Accepted 9 August 2017
Available online 10 August 2017

Keywords:
SIR infection model
Stochastic model
Markov decision theory
Epidemic control

ABSTRACT

The emergence of a new strain of a disease, or the introduction of an existing strain to a naive population, can give rise to an epidemic. We consider how to maximise the probability of epidemic fade-out – that is, disease elimination in the trough between the first and second waves of infection – in the Markovian SIR-with-demography epidemic model. We assume we have an intervention at our disposal that results in a lowering of the transmission rate parameter, β, and that an epidemic has commenced. We determine the optimal stage during the epidemic in which to implement this intervention. This may be determined using Markov decision theory, but this is not always practical, in particular if the population size is large. Hence, we also derive a formula that gives an almost optimal solution, based upon the approximate deterministic behaviour of the model. This formula is explicit, simple, and, perhaps surprisingly, independent of β and the effectiveness of the intervention. We demonstrate that this policy can give a substantial increase in the probability of epidemic fade-out, and we also show that it is relatively robust to a less than ideal implementation.

1. Introduction

One of the key goals of epidemiology is to take action to minimise the impact of epidemic outbreaks. With this in mind, many studies have investigated ways to optimise the control of an outbreak. Good overviews can be found in Kar and Batabyal [12] and Yaesoubi and Cohen [28].

Studies tend to take one of two approaches: investigating either the use of vaccination [25] (reducing the susceptible population), or prophylactic measures to reduce the spread of the infection [22]. Prophylactic measures include antivirals [4,9,17,18], or non-pharmaceutical interventions such as school closures [10] or the teaching of basic personal health habits [21]. Vaccination is usually the ideal, but it is often not available in the early stages of a novel strain/disease. Antivirals, or other measures to reduce the spread of an infection, are therefore an important tool in attempting to control an outbreak.

Previous studies have concentrated on either the initial stages of an infection – and measures to prevent the infection becoming an outbreak – or an established infection, endemic to a population. In this paper, we instead examine *epidemic fade-out*, which has been nominated as an area requiring more research [5,6], and has not previously been investigated in terms of control.

Epidemic fade-out refers to the case in which an infection has a large initial outbreak, and it is eliminated from the population in the first trough after that initial outbreak [19]. Therefore, techniques to maximise the probability of epidemic fade-out offer the opportunity to prevent an infection from becoming established – that is, endemic – in a population.

We use the Markovian SIR-with-demography infection model [20]. Important previous work was by van Herwaarden [26] and Meerson and Sasorov [19], who both provided methods for approximating the probability of epidemic fade-out for this model. van Herwaarden used the Fokker-Plank approximation, while Meerson and Sasorov used the WKB appoximation. Both of these papers gave explicit formulae for the probability of epidemic fade-out, to a good degree of accuracy. In a previous paper [3] we outlined a more accurate numerical approximation method, and also presented a range of results from our calculations. These results showed that the probability of epidemic fade-out is non-monotonic in the transmission rate parameter β. Typically, a lower value of β increases the probability of epidemic fade-out, which is the intuitive result (less transmission $\rightarrow$ higher probability of fade-out). But in some situations, perhaps counter-intuitively, a reduction in the value of β causes the probability of epidemic fade-out to decrease.

Similar examples of non-monotonicity in epidemics – of a reduction in transmission or an increase in treatment actually increasing the total epidemic size, or making the epidemic more likely to persist – have been reported by others, but in different

* Corresponding author.
E-mail addresses: peter.ballard@adelaide.edu.au (P.G. Ballard), nigel.bean@adelaide.edu.au (N.G. Bean), joshua.ross@adelaide.edu.au (J.V. Ross).

http://dx.doi.org/10.1016/j.mbs.2017.08.003
0025-5564/

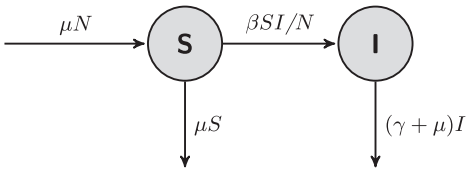


Fig. 1. The SIR-with-demography epidemic model. S is the number of susceptibles and I is the number of infectious individuals. N is a fixed parameter, but the population size is not fixed. β is the transmission rate parameter, μ is the per-capita birth/death rate, and γ is the recovery rate.

Table 1
Transition rates for the Markovian SIR-with-demography epidemic model displayed in Fig. 1.

Description	Transition	Rate
Infection	$(S, I) \to (S-1, I+1)$	$\beta SI/N$
Birth of susceptible	$(S, I) \to (S+1, I)$	μN
Death of susceptible	$(S, I) \to (S-1, I)$	μS
Removal of infectious	$(S, I) \to (S, I-1)$	$(\gamma+\mu)I$

contexts. Feng et al. [7], Rozhnova et al. [24], and Lee and Chowell [15] all reported non-monotonicity in the context of seasonal forcing. Xiao et al. [27] saw it in the case of multiple strains of an infection. Grigorieva and Khailov [8] is perhaps the closest analogue to this paper. In a deterministic SIR model, they showed that not reducing β early in the infection cycle can minimise the total epidemic size.

For epidemic fade-out, the non-monotonicity in the transmission rate parameter β suggests that there are two or more competing effects, and that in some states a higher β will maximise the probability of epidemic fade-out, and in other states a lower β will maximise it. So it should be possible to find the optimal policy for choosing higher or lower β. Finding this optimal policy is the topic of this paper.

We will show that this optimal policy entails delaying the implementation of measures to reduce β, resulting in more individuals being infected in the short term. In lethal epidemics, even if the long term result is a more likely fade-out and hence the elimination of the disease from the population, this is likely to be impossible to do ethically. Therefore, the applicability of this method will probably be limited to situations of non-lethal infections, or diseases among animals.

We examine two different control scenarios: an idealised scenario in Section 3 and a more realistic scenario in Section 4. Section 3.3 is the most significant contribution of the paper, where we derive a simple control policy that is a close approximation to the optimal control policy in the idealised scenario. Section 4.3 supplements Section 3.3, by showing that the same simple control policy is also a close approximation to the optimal control policy in the realistic scenario. Effectively, we provide an explicit, simple rule for when to implement an intervention. Perhaps surprisingly, this rule is independent of the transmission rate parameter β and the effectiveness of the intervention. The results, which show a significant increase in the probability of epidemic fade-out when using any of these methods, are presented in Section 5.

2. Model and definitions

2.1. The SIR-with-demography model

We use the Markovian SIR-with-demography model, as described in Fig. 1 and Table 1. S and I represent the number of "susceptible" and "infectious" individuals respectively. The parameters β, γ and μ are all strictly positive. The number of "recovered" individuals (R) is usually included in the model, but is redundant and can be removed from the analysis by considering "death of infectious" (at rate μI) and "recovery of infectious" (at rate γI) as equivalent [11]. We use a common death rate μ, corresponding to a non-lethal infection, as this is the original and most common model [1]. If a different death rate μ_I is used for infectious individuals, then the analysis in the rest of this paper follows similarly, if one replaces all references to $(\gamma+\mu)$ with $(\gamma+\mu_I)$.

In the limit as N becomes large, a suitably scaled version of the stochastic process converges (uniformly in probability over finite time intervals) to a deterministic process [14]; this provides an approximation to the expected dynamics, for finite N, governed by the differential equations:

$$\frac{dS}{dt} = \mu(N-S) - \beta SI/N,$$
$$\frac{dI}{dt} = \beta SI/N - (\gamma+\mu)I. \tag{1}$$

We refer to this as the *deterministic approximation*.

In a naive population, $S \approx N$. So R_0, the basic reproduction number, is given by:

$$R_0 = \frac{\beta}{\gamma+\mu}. \tag{2}$$

We are only concerned with cases in which $R_0 > 1$, when a major outbreak may occur. In these cases, the endemic point is where both derivatives in (1) are equal to zero, and is given by:

$$(S_e, I_e) = N\left(\frac{\gamma+\mu}{\beta}, \frac{\mu(\beta-\gamma-\mu)}{\beta(\gamma+\mu)}\right). \tag{3}$$

The stability analysis of (1), linearised around the endemic point (S_e, I_e), determines that the eigenvalues λ obey,

$$\lambda^2 + R_0\mu\lambda + \mu(\gamma+\mu)(R_0-1) = 0. \tag{4}$$

The trajectory of the deterministic approximation is oscillatory if and only if these eigenvalues are complex [13], which in turn requires,

$$\frac{\gamma+\mu}{\mu} > \frac{{R_0}^2}{4(R_0-1)}. \tag{5}$$

In any realistic system, $\gamma \gg \mu$ and inequality (5) is comfortably met. In that case, the trajectory of the deterministic approximation of a typical outbreak is shown in Fig. 2. It starts at point A, rises to a peak (B), falls through C to a first local minimum (D), and converges in a spiral towards the endemic point.

Since the stochastic model has discrete states, it is sometimes convenient to round the endemic state values up to the next highest integer pair:

$$(S_d, I_d) = (\lceil S_e \rceil, \lceil I_e \rceil). \tag{6}$$

It can also be shown from (1) and (3) that dI/dt is positive for $S > S_e$ and negative for $S < S_e$, that is:

$$\operatorname{sgn}\left(\frac{dI}{dt}\right) = \operatorname{sgn}(S - S_e). \tag{7}$$

As we mentioned above, the expected behaviour of the CTMC (continuous-time Markov chain) tracks the deterministic approximation as N becomes large. However, a stochastic realisation may fade out at the start (near point A), or in the first trough after the initial outbreak (near point D). The latter situation, known as epidemic fade-out, is the topic of this paper.

The initial state (point A) is (S_0, I_0). In all our calculations, I_0 is small and $S_0 = N - I_0$; this represents the beginning of an outbreak in a naive population.

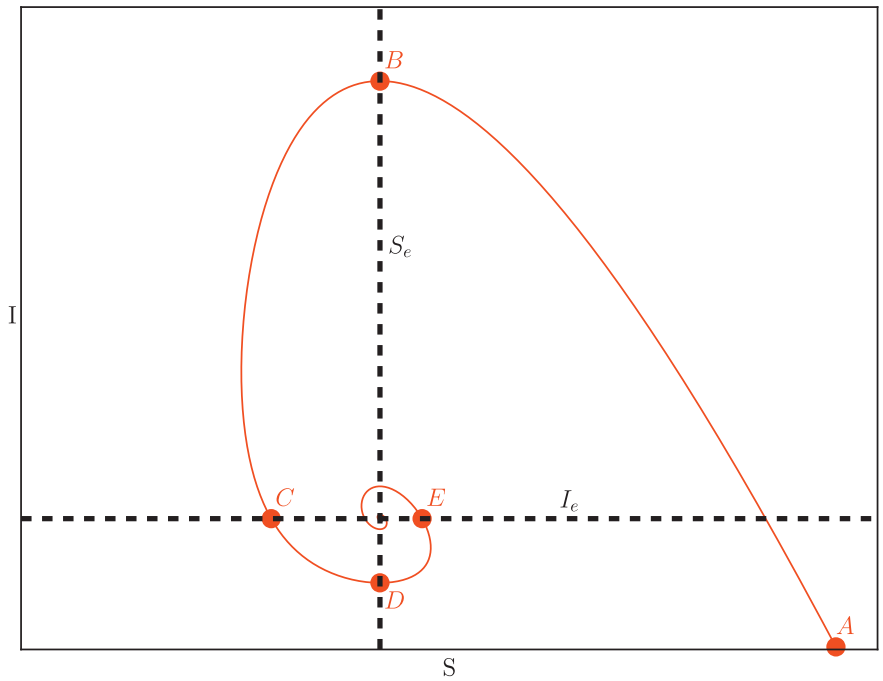


Fig. 2. I versus S plot of the deterministic approximation of a typical SIR-with-demography model. It starts at point A, rises to B, then falls through C to D before rising to E and eventually converging on the endemic point (S_e, I_e). An actual stochastic realisation may fade out to $I = 0$ near point D, an effect known as epidemic fade-out. It follows from (7) that points B and D are both at $S = S_e$.

2.2. Different transmission rate parameters

During the initial outbreak, a higher value of β causes S to fall to a lower value, which in turn can cause I to fall to a lower value during the first trough, increasing the probability of epidemic fade-out. On the other hand, when the first wave of infection subsides and the number of infectious individuals becomes very low, a lower value of β causes an increase in the probability of epidemic fade-out.

Therefore, we conjecture that the optimal strategy is to allow a higher β early in the outbreak, and to implement the lower β later in the outbreak, as the CTMC approaches the first trough. (This is confirmed in the results in Section 5.1).

The initial transmission rate parameter, corresponding to intervention measures not being in place, is denoted $\beta^{(1)}$, and the value whilst the intervention is implemented is denoted $\beta^{(2)}$. We specify that $\beta^{(1)} > \beta^{(2)}$.

We also use a superscript in parentheses to represent variables corresponding to the use of $\beta^{(1)}$ or $\beta^{(2)}$; so for instance $P^{(k)}$ is the transition probability matrix when using $\beta = \beta^{(k)}$, for $k = 1, 2$.

2.3. Definition of epidemic fade-out

Informally, epidemic fade-out refers to fade-out during the first trough after the initial substantial wave of infection, roughly between points C and E in Fig. 2. But for calculations, it is important to have a precise definition. (In general the exact definition is not overly critical, as long as it is used consistently). To do so, we need to define the pre-condition (that a substantial outbreak has commenced), and then need to define what constitutes fade-out in the first trough (or conversely, what constitutes an escape from the first trough).

We define $\mathbb{S}$ to be the state space of all possible (S, I) values, and we define T to be:

$$T = \{(S, I) \in \mathbb{S} | S = S_d^{(1)} - 1\}. \tag{8}$$

T is illustrated by the green dotted line in Fig. 3. We define that the initial wave of infection has occurred if the CTMC reaches a state in T. We use this definition because if the CTMC satisfies this condition, it can be called a substantial outbreak, so a fade-out in the subsequent trough can reasonably be called epidemic fade-out.

Given that the CTMC reaches T, it will almost surely eventually fall to a state for which $I < I_d^{(1)}$. Therefore we define a two boundary hitting problem: epidemic fade-out occurs if the CTMC reaches a lower absorbing boundary L, before it reaches an artificial upper absorbing boundary U. The lower absorbing boundary is

$$L = \{(S, I) \in \mathbb{S} | I = 0\}.$$

In a CTMC in which β is constant, the deterministic approximation of the CTMC converges to a point near (S_d, I_d), as given by (6), taking an anticlockwise path in Fig. 2. So in that case, the line $U = \{(S, I) \in \mathbb{S} | S \geq S_d, I = I_d\}$ would be a natural definition of the upper absorbing boundary [3,26].

However in a CTMC in which β can take two different values, there are two possible solutions to (6), depending on the value of β: $\left(S_d^{(1)}, I_d^{(1)}\right)$ (given by setting $\beta = \beta^{(1)}$ in (3)); and $\left(S_d^{(2)}, I_d^{(2)}\right)$ (given by setting $\beta = \beta^{(2)}$ in (3)); where $S_d^{(1)} \leq S_d^{(2)}$ and $I_d^{(1)} \geq I_d^{(2)}$.

To account for the possibility that β may be either $\beta^{(1)}$ or $\beta^{(2)}$, we end the first trough at $I = I_d^{(1)}$ for $S_d^{(1)} \leq S < S_d^{(2)}$, and at $I = I_d^{(2)}$ for $S_d^{(2)} \leq S$. We join the two boundaries with a vertical boundary

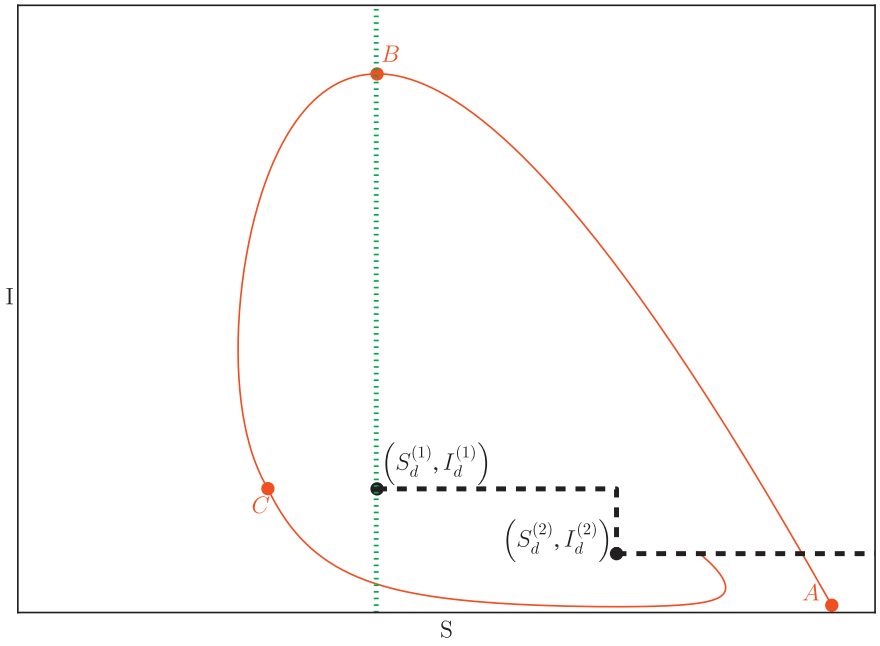


Fig. 3. The probability of epidemic fade-out is the probability of the CTMC being absorbed at L ($I = 0$) before next reaching U (the black (dashed) line), given that it reaches T (the green (vertical dotted) line). The red (solid) line shows the behaviour of the deterministic approximation to the CTMC, starting at point A; with the transmission rate parameter changing from $\beta^{(1)}$ to $\beta^{(2)}$ at point C.

at $S = S_d^{(2)}$ for $I_d^{(2)} \le I \le I_d^{(1)}$, giving:

$$
\begin{aligned}
U &= U_1 \cup U_2 \cup U_3, \text{ where} \\
U_1 &= \{(S, I) \in \mathbb{S} | I = I_d^{(1)}, S_d^{(1)} \le S \le S_d^{(2)}\}, \\
U_2 &= \{(S, I) \in \mathbb{S} | S = S_d^{(2)}, I_d^{(2)} \le I \le I_d^{(1)}\}, \text{ and} \\
U_3 &= \{(S, I) \in \mathbb{S} | I = I_d^{(2)}, S_d^{(2)} \le S\}.
\end{aligned}
$$

This is illustrated in Fig. 3. We therefore define, p_0, the probability of epidemic fade-out, as the probability that the process is absorbed at L before reaching a state in U, given that it reaches T.

3. Idealised scenario: activation and de-activation of $\beta^{(2)}$

We first consider an idealised scenario, in which it is possible to switch an unlimited number of times between using $\beta^{(1)}$ and $\beta^{(2)}$, corresponding to activation and de-activation of the intervention measure, in a state dependent manner.

3.1. Definition of the problem

To give a practical solution, we limit the state space to be finite. So for the (infinite) state space $\mathbb{S}$, we define $\mathbb{S}'$ as the finite set of states,

$$\mathbb{S}' = \{(S, I) | 0 \le S \le (1.1)N, 0 \le I \le (1.1)N\}. \tag{9}$$

We enforce this by modifying the CTMC so that in Table 1, "Infection" events may not occur if $I \ge (1.1)N$ and "Birth of susceptible" events may not occur if $S \ge (1.1)N$. In all but the smallest systems, reaching these states is extremely improbable, so this modification has a negligible impact on the results.

We also define M as the set of transient states in $\mathbb{S}'$, that is all states in neither absorbing boundary:

$$M = \mathbb{S}' \setminus (U \cup L). \tag{10}$$

Then for some set $V \subseteq M$, the policy is:

$$\beta = \begin{cases} \beta^{(2)} & \text{if } (S, I) \in V \text{ and the CTMC has previously reached } T; \\ \beta^{(1)} & \text{otherwise.} \end{cases}$$

Problem. Find the set V which maximises p_0.

3.2. Optimal solution

This idealised scenario can be regarded as a Markov decision process, with an infinite horizon and no discounting, in which the "reward" is gained by reaching absorption at L. It can be solved by using a policy iteration algorithm [23, Section 7.2.5].

We define two transition probability matrices, $P^{(1)}$ and $P^{(2)}$ on M, corresponding to the jump chain of the CTMC [2]. The transition probabilities are calculated from the rates specified in Table 1; where $P_{ij}^{(k)}$ is the transition probability from state i to state j when using $\beta^{(k)}$ as the transmission rate parameter, for $k = 1, 2$. Transitions to absorbing states are not included, so some rows will sum to less than 1. These matrices are sparse, with at most four non-zero entries in each row.

Since we do not include the absorbing states in $P^{(1)}$ and $P^{(2)}$, the "reward" is only earned from states adjacent to the $I = 0$ boundary. We set up the respective reward vectors, $R^{(1)}$ and $R^{(2)}$, in which the reward is the probability of being absorbed at $I = 0$ on the next step. (Hence values will be non-zero only for states with $I = 1$.)

We create a decision vector $D(n)$ for each policy iteration step n. It has one entry per state in M, and every entry must be either 1 or 2. For each step n we also create the matrix $P(n)$ and the vector $R(n)$ according to the rule,

$$
\begin{aligned}
P_{ij}(n) &= P_{ij}^{(D_i(n))}, \\
R_i(n) &= R_i^{(D_i(n))},
\end{aligned} \tag{11}
$$

for all $i, j \in M$.

The goal is to find the vector $D(n)$ which maximises p_0. Then V is the set of states $i \in M$ for which $D_i(n) = 2$.

The policy iteration algorithm solves this problem as follows:

1. Set $n = 1$ and initialise $D(n)$ to any permissible vector.
2. Build $P(n)$ and $R(n)$, according to (11).
3. Determine the column vector $v(n)$ by solving:

$$(\mathbb{I} - P(n))v(n) = R(n), \tag{12}$$

where $\mathbb{I}$ is the identity matrix. Then for each state $i \in M$, let

$$v_i^{(1)}(n) = R_i^{(1)} + \sum_{j\in M} P_{ij}^{(1)} v_j(n), \tag{13}$$

$$v_i^{(2)}(n) = R_i^{(2)} + \sum_{j\in M} P_{ij}^{(2)} v_j(n), \tag{14}$$

$$z_i(n) = \operatorname{sgn}\left(v_i^{(2)}(n) - v_i^{(1)}(n)\right). \tag{15}$$

(We use $v_i^{(1)}(n)$ and $v_i^{(2)}(n)$ on the left hand side of (13) and (14) because these equations are re-calculating $v_i(n)$ assuming $D_i(n) = 1$ and $D_i(n) = 2$ respectively, with the rest of the elements in $D(n)$ unchanged).

4. Update the policy: for each state i in D,

$$D_i(n+1) = \begin{cases} 1 & \text{if } z_i(n) = -1, \\ 2 & \text{if } z_i(n) = 1, \\ D_i(n) & \text{if } z_i(n) = 0. \end{cases} \tag{16}$$

5. If $D(n+1) = D(n)$, then the algorithm terminates, and V is the set of states i for which $D_i(n) = 2$. Otherwise, increment n and repeat from Step 2.

Notice that $v_i(n)$ is the probability of hitting L before U (for transition matrix $P(n)$ and reward vector $R(n)$), given that the CTMC is in state i. The policy iteration algorithm finds the $P(n)$ and $R(n)$ which give the maximum $v_i(n)$ for all $i \in M$ [23, Proposition 7.2.14], and hence finds the policy which gives the maximum p_0, regardless of the initial state (S_0, I_0).

3.3. Simplifying the policy iteration algorithm

The technique in Section 3.2 is useful for small populations, but becomes impractical for even moderate population sizes. For instance, a population with $N = 1000$ would have approximately 10^6 states in M. The corresponding two-dimensional matrix $(\mathbb{I} - P(n))$ in (12) is then approximately $10^6 \times 10^6$, and even though it is sparse, solving (12) takes significant computing resources. Therefore, it would be beneficial to find a solution method which avoids the need to solve (12).

Let q_{ij} be the transition rate between any two states i and j, $i \neq j$, if $\beta^{(1)}$ is the transmission rate parameter (where $i \in M$ and $j \in \mathbb{S}'$). For notational convenience we let $q_{ii} = 0$. Also let $q_i = \sum_{j\in\mathbb{S}'} q_{ij}$ be the sum of all transition rates out of state i when $\beta^{(1)}$ is the transmission rate parameter. So $P_{ij}^{(1)} = \frac{q_{ij}}{q_i}$ and $R_i^{(1)} = \frac{\sum_{j\in L} q_{ij}}{q_i}$. In that case we may rewrite (13) as,

$$v_i^{(1)}(n) = \sum_{j\in L}\left(\frac{q_{ij}}{q_i}\right) + \sum_{j\in M}\left(\frac{q_{ij}}{q_i}\right) v_j(n).$$

In order to unify these two sums, we also define $v_j(n) = 1$ for $j \in L$, and $v_j(n) = 0$ for $j \in U$. Then,

$$v_i^{(1)}(n) = \frac{\sum_{j\in\mathbb{S}'} q_{ij} v_j(n)}{q_i}. \tag{17}$$

For every state $i \in M$, except those for which $S_i = 0$, let h be the state that is reached from i by an infection event, and let $\delta_i = (\beta^{(1)} - \beta^{(2)}) S_i I_i / N$, where S_i and I_i are the S and I values corresponding to state i. That is, $q_{ih} - \delta_i$ is the transition rate from state i to state h when using $\beta^{(2)}$ as the transmission rate parameter. (The $S_i = 0$ case is excluded because in that case no infection event is possible, so $v_i^{(1)}(n) = v_i^{(2)}(n)$, so (15) always evaluates to zero and it never matters whether or not i is in V.) Thus, $S_i > 0$ ensures that $\delta_i > 0$. Then (14) becomes:

$$v_i^{(2)}(n) = \frac{\sum_{\{j\in\mathbb{S}', j\neq h\}} q_{ij} v_j(n)}{q_i - \delta_i} + \frac{(q_{ih} - \delta_i) v_h(n)}{q_i - \delta_i}$$

$$\Rightarrow v_i^{(2)}(n) = \frac{\sum_{j\in\mathbb{S}'} q_{ij} v_j(n)}{q_i - \delta_i} - \frac{\delta_i v_h(n)}{q_i - \delta_i}. \tag{18}$$

Substituting in (17) gives,

$$v_i^{(2)}(n) = \left(\frac{q_i}{q_i - \delta_i}\right) v_i^{(1)}(n) - \frac{\delta_i v_h(n)}{q_i - \delta_i}$$

$$\Rightarrow v_i^{(2)}(n) - v_i^{(1)}(n) = \left(\frac{q_i}{q_i - \delta_i} - 1\right) v_i^{(1)}(n) - \frac{\delta_i v_h(n)}{q_i - \delta_i}$$

$$\Rightarrow v_i^{(2)}(n) - v_i^{(1)}(n) = \frac{\delta_i\left(v_i^{(1)}(n) - v_h(n)\right)}{q_i - \delta_i}. \tag{19}$$

Then noting that $\delta_i > 0$ and $q_i - \delta_i > 0$,

$$\operatorname{sgn}\left(v_i^{(2)}(n) - v_i^{(1)}(n)\right) = \operatorname{sgn}\left(v_i^{(1)}(n) - v_h(n)\right). \tag{20}$$

We can also substitute (17) and then (18) into the right hand side of (19), to give:

$$v_i^{(2)}(n) - v_i^{(1)}(n) = \frac{\delta_i}{q_i - \delta_i}\left(\frac{\sum_{j\in\mathbb{S}'} q_{ij} v_j(n)}{q_i} - v_h(n)\right)$$

$$= \frac{\delta_i}{q_i - \delta_i}\left(\frac{\sum_{j\in\mathbb{S}'} q_{ij} v_j(n)}{q_i} - \frac{\delta_i v_h(n)}{q_i} - \frac{(q_i - \delta_i) v_h(n)}{q_i}\right)$$

$$= \frac{\delta_i}{q_i - \delta_i}\left(\frac{(q - \delta_i) v_i^{(2)}(n)}{q_i} - \frac{(q_i - \delta_i) v_h(n)}{q_i}\right)$$

$$\Rightarrow v_i^{(2)}(n) - v_i^{(1)}(n) = \frac{\delta_i\left(v_i^{(2)}(n) - v_h(n)\right)}{q_i}$$

$$\Rightarrow \operatorname{sgn}\left(v_i^{(2)}(n) - v_i^{(1)}(n)\right) = \operatorname{sgn}\left(v_i^{(2)}(n) - v_h(n)\right). \tag{21}$$

Comparing (12) to (13) and (14) tells us that $v_i(n)$ is equal to either $v_i^{(1)}(n)$ or $v_i^{(2)}(n)$, so (20) and (21) combine to give,

$$\operatorname{sgn}\left(v_i^{(2)}(n) - v_i^{(1)}(n)\right) = \operatorname{sgn}(v_i(n) - v_h(n)),$$

which can then be substituted into (15). So (12)–(14) can be removed from Step 3 of the policy iteration algorithm, which simplifies to:

3. For each state i,

$$z_i(n) = \operatorname{sgn}(v_i(n) - v_h(n)). \tag{22}$$

The meaning of (22) and (16) is that we should reduce the transition rate from i to h only if $v_i(n) > v_h(n)$, which is a reasonably intuitive result.

An important feature of (22) is that it does not necessarily include a full matrix calculation. This opens the possibility of simpler ways to calculate an optimal, or close to optimal, policy. For instance, (22) could be evaluated for a small number of states, using an approximate method as in [3] to calculate $v_i(n)$ and $v_h(n)$.

3.3.1. A simplified policy based on the deterministic local minimum

A further advantage of (22) is that we do not need to calculate $v_i(n)$ and $v_h(n)$ at all. We only need to calculate which is greater.

We can get a very quick approximation of $\mathrm{sgn}(v_i(n) - v_h(n))$ in (22) by taking advantage of a property which we reported previously [3]: for a state x in the region where dI/dt of the deterministic approximation is negative or zero (which means $S \le S_e^{(1)}$, by (7)), $v_x(n)$ is generally negatively correlated to the minimum I value of the deterministic curve beginning at x.

That is, to a good approximation, the closer the curve of the deterministic approximation comes to an absorbing boundary, the more likely the process is to be absorbed at that boundary. A deterministic curve with a lower minimum passes closer to the absorbing boundary, and is closer to that boundary for a longer time; both of these effects contribute to making epidemic fade-out more probable.

Furthermore, it is possible to reduce this comparison to a formula. It is preferable to reduce β when $v_h < v_i$, which means (by our approximation) that h must be on a "higher" deterministic curve than i. A line from i to h (that is, from (S, I) to $(S-1, I+1)$) has a slope of -1. Since $dI/dt < 0$ in this region, a step of slope -1 goes to a "higher" deterministic curve when $dS/dt > -dI/dt$; that is, if:

$$(\gamma + \mu)I < \mu(N - S). \tag{23}$$

For states in the region where dI/dt of the deterministic approximation is positive (which means at least for $S > S_e^{(2)}$), we cannot use this approximation because the deterministic minimum has already been passed. However, h is always on a higher (further from $I = 0$) curve than i, as well as having a higher I value; so $\beta = \beta^{(2)}$ is always preferred.

In the region $S_e^{(1)} < S \le S_e^{(2)}$: if at any point we "try" $\beta = \beta^{(1)}$, this means that $S_e = S_e^{(1)}$, so $S > S_e$. Then we find (by the analysis in the previous paragraph) that $\beta = \beta^{(2)}$ is preferred. So this means that $\beta = \beta^{(2)}$ is preferred for all $S > S_e^{(1)}$.

Putting this together gives the following set V, the set of states in which to use $\beta^{(2)}$:

$$\begin{aligned} V &= V_1 \cup V_2 \cup V_3, \text{ where} \\ V_1 &= \{(S,I) \in \mathbb{S} | S \le S_e^{(1)}, (\gamma+\mu)I < \mu(N-S)\}, \\ V_2 &= \{(S,I) \in \mathbb{S} | S_e^{(1)} < S \le S_e^{(2)}, I < I_e^{(1)}\}, \text{ and} \\ V_3 &= \{(S,I) \in \mathbb{S} | S_e^{(2)} < S, I < I_e^{(2)}\}. \end{aligned} \tag{24}$$

This formula is explicit, and it is quick to calculate regardless of the population size. It specifies V with a simple line, as illustrated by the green (top) line in Fig. 4.

If the CTMC roughly follows the deterministic approximation (Figs. 2 and 3) then V is first entered when $S < S_e^{(1)}$. So the most important component of (24) is V_1, as specified in (23). Note, importantly and possibly surprisingly, that (23) is independent of the values of $\beta^{(1)}$ and $\beta^{(2)}$. So the condition for initially using $\beta^{(2)}$ does not depend on the values of $\beta^{(1)}$ and $\beta^{(2)}$.

The formula also tells us, at least assuming the approximation used here, that the optimal policy cannot be improved by allowing three or more values of β. The optimum is always to use the highest available β for states not in V (corresponding to no intervention), and the lowest available β (corresponding to the most effective set of interventions) for states in V.

4. Realistic scenario: activation only of $\beta^{(2)}$

4.1. Definition of the problem

The idealised scenario in Section 3 corresponds to the most effective intervention possible, but it is not realistic. It allows the CTMC to repeatedly switch between using $\beta^{(1)}$ and $\beta^{(2)}$ as the state changes. In most real-world situations it would not be practical to start and stop infection-reducing measures as the process changes state near the boundary of V.

A more realistic situation, which we refer to as the *realistic scenario*, allows activation of $\beta^{(2)}$ only once. In this scenario, once $\beta = \beta^{(2)}$ is used, $\beta = \beta^{(2)}$ is *always* used (until the boundary L or U is reached), even if the CTMC subsequently leaves the region V. This is more practical because, in a typical application, measures to reduce the infection rate would be kept in place for a reasonable length of time once they are implemented.

As in Section 3.2, T is defined in (8), and M is defined in (10). Then for some set $V \subseteq M$, the policy is:

Initially, $\beta = \beta^{(1)}$. When the CTMC reaches a state in V, having previously been in a state in T, it permanently uses $\beta = \beta^{(2)}$.

Problem. Find the set V which maximises p_0.

4.2. Optimal solution

Since it is a relatively simple task to calculate the absorption probability once we are permanently using $\beta = \beta^{(2)}$, the realistic scenario can be regarded as an "optimal stopping" problem. The optimal stopping algorithm is as follows [23, Section 7.2.8]:

1. Build $P^{(1)}$, $R^{(1)}$, $P^{(2)}$ and $R^{(2)}$ as in Section 3.2.
2. Find $v^{(2)}$, the solution to
$$\left(\mathbb{I} - P^{(2)}\right)v^{(2)} = R^{(2)}. \tag{25}$$
Now, $v^{(2)}$ is the vector of absorption probabilities assuming the transmission rate parameter is fixed at $\beta^{(2)}$, which form the "stopping rewards". So in state i, the CTMC can "stop" (switch to $\beta = \beta^{(2)}$) and take the "stopping reward" $v_i^{(2)}$.
3. Find, by linear programming, the vector v with the minimum $\Sigma_i v_i$, subject to the constraints:
$$v_i \ge \sum_{j \in M} P_{ij}^{(1)} v_j + R_i^{(1)} \quad \text{and} \quad v_i \ge v_i^{(2)}, \quad \forall i \in M. \tag{26}$$
4. Create V to represent the optimal policy. For each state i, if $v_i = v_i^{(2)}$, then the optimal policy in state i is to switch to using $\beta^{(2)}$, so i is added to V. If $v_i > v_i^{(2)}$, then the optimal policy in state i is to continue using $\beta^{(1)}$, so i is not added to V.

As in Section 3.2, v_i is the probability of hitting L before U, given that the CTMC is in state i. The vector $v = (v_i, i \in M)$, can also be calculated from a given V without running the algorithm: row i of P is zero if $i \in V$, and is otherwise equal to row i of $P^{(1)}$; element i of R is equal to $v_i^{(2)}$ if $i \in V$, and is otherwise equal to $R_i^{(1)}$; and v is the solution to,

$$(\mathbb{I} - P)v = R. \tag{27}$$

4.3. Simplifying the optimal stopping algorithm

If we use the same definitions for q_{ij}, q_i, δ_i and h as in Section 3.3, then the solution to (25) and (26) satisfies:

$$v_i^{(1)} = \frac{\sum_{j \in \mathbb{S}'} q_{ij} v_j}{q_i}, \tag{28}$$

$$v_i^{(2)} = \frac{\sum_{j \in \mathbb{S}'} q_{ij} v_j^{(2)}}{q_i - \delta_i} - \frac{\delta_i v_h^{(2)}}{q_i - \delta_i}, \quad \text{and} \tag{29}$$

$$v_i = \max\left(v_i^{(1)}, v_i^{(2)}\right), \tag{30}$$

for all $i \in M$.

Another way of expressing (30) is to say,

$$z_i = \mathrm{sgn}\left(v_i^{(2)} - v_i^{(1)}\right); \tag{31}$$

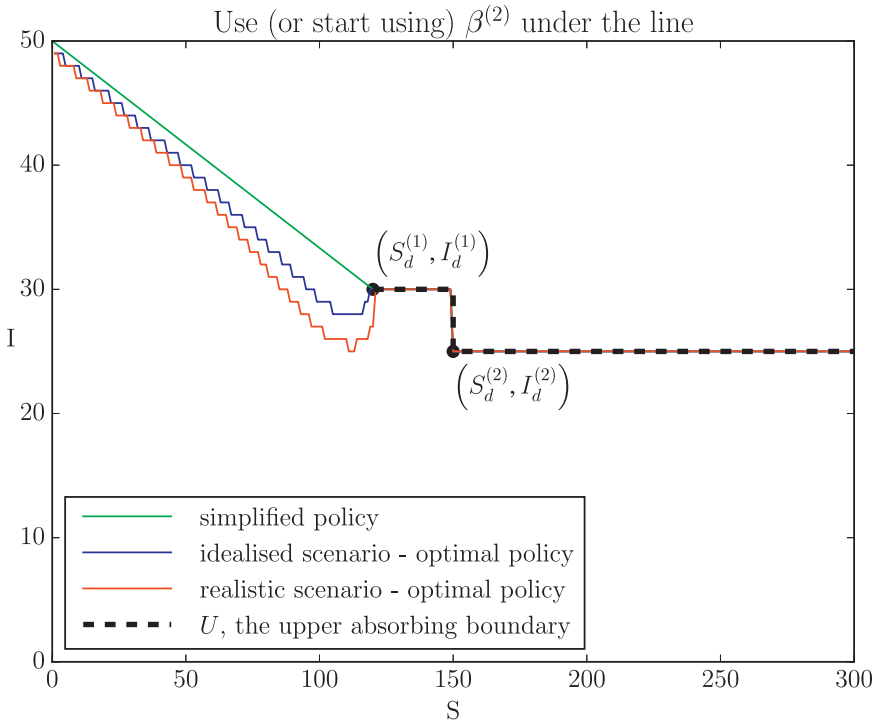


Fig. 4. Comparison of the policies for exact methods of Sections 3.2 and 4.2, and the simplified method of Section 3.3.1. The parameters are $N = 300$, $\beta^{(1)} = 3$, $\beta^{(2)} = 2.4$, $\gamma = 1$ and $\mu = 0.2$. In each scenario, the policy (V) is the set of states below the respective line. Lines are in the same vertical order as the legend box. All lines are coincident along U. The endemic points corresponding to $\beta^{(1)}$ and $\beta^{(2)}$, $\left(S_d^{(1)}, I_d^{(1)}\right)$ and $\left(S_d^{(2)}, I_d^{(2)}\right)$ respectively, are marked.

where $v_i = v_i^{(2)}$ if $z_i = 1$, $v_i = v_i^{(1)}$ if $z_i = -1$, and v_i may be either if $z_i = 0$. In that case, we see that the (28), (29) and (31) are identical to (17), (18) and (15) respectively with the "(n)" postscripts removed, with the exception of the use of $v_i^{(2)}$ instead of v_i on the right hand side of (29). Then,

$$v_i^{(2)} - v_i^{(1)} = \frac{\sum_{j\in\mathbb{S}'} q_{ij} v_j^{(2)}}{q_i - \delta_i} - \frac{\delta_i v_h^{(2)}}{q_i - \delta_i} - v_i^{(1)}$$

$$= \frac{\sum_{j\in\mathbb{S}'} q_{ij}\left(v_j^{(2)} - v_j\right)}{q_i - \delta_i} + \frac{\sum_{j\in\mathbb{S}'} q_{ij} v_j}{q_i - \delta_i} - \frac{\delta_i v_h^{(2)}}{q_i - \delta_i} - v_i^{(1)}$$

$$= \frac{\sum_{j\in\mathbb{S}'} q_{ij}\left(v_j^{(2)} - v_j\right)}{q_i - \delta_i} + \frac{q_i v_i^{(1)}}{q_i - \delta_i} - \frac{\delta_i v_h^{(2)}}{q_i - \delta_i} - v_i^{(1)}$$

$$= \frac{\sum_{j\in\mathbb{S}'} q_{ij}\left(v_j^{(2)} - v_j\right)}{q_i - \delta_i} + \frac{\delta_i\left(v_i^{(1)} - v_h^{(2)}\right)}{q_i - \delta_i}$$

$$= \frac{\delta_i\left(v_i^{(1)} - v_h\right)}{q_i - \delta_i} + \frac{\left(\sum_{j\in\mathbb{S}'} q_{ij}\left(v_j^{(2)} - v_j\right)\right) - \delta_i\left(v_h^{(2)} - v_h\right)}{q_i - \delta_i}$$

$$\Rightarrow z_i = \operatorname{sgn}\left(\delta_i\left(v_i^{(1)} - v_h\right) + \left[\left(\sum_{j\in\mathbb{S}'} q_{ij}\left(v_j^{(2)} - v_j\right)\right) - \delta_i\left(v_h^{(2)} - v_h\right)\right]\right). \tag{32}$$

Although it may not be immediately obvious, (32) is quite similar to (22). The only differences are: the expression in square brackets; the presence of $v_i^{(1)}$ instead of v_i; and the inclusion of δ_i. Note that since $v_i \geq v_i^{(1)}$ and $v_i \geq v_i^{(2)}$ for all i, the right hand side of (32) cannot be greater than the right hand side of (22), so the criterion for including a state i in V is always more stringent in the realistic scenario than in the idealised scenario.

The presence of $v_i^{(1)}$ is due to the fact that the change is one-way, so for the initial application of the policy (that is, for the first iterative change from $v_i = v_i^{(1)}$ to $v_i = v_i^{(2)}$ in Section 3.2) the two equations are identical if the expression in square brackets is zero. The expression in square brackets accounts for whether the states surrounding state i are in V, and the q_{ij} and δ_i terms act as weighting factors.

However, when we look at the actual optimal policies generated by (22) (Fig. (4)), we see that most states in V are surrounded by other states in V, making the expression in square brackets equal to zero. So that suggests that (32) will produce a V very similar to the V corresponding to the optimal policy for the realistic scenario – that is, that the optimal policies for the two scenarios will have very similar sets V.

This in turn suggests that using the set V defined in (24) will also be a good approximation of the optimal policy for the realistic scenario, as we see in the following section.

5. Results

We refer to the policy calculated in Section 3.2 as the *idealised scenario – optimal policy*, the policy calculated in Section 4.2 as the *realistic scenario – optimal policy*, and the policy calculated using (24) in Section 3.3.1 as the *simplified policy*.

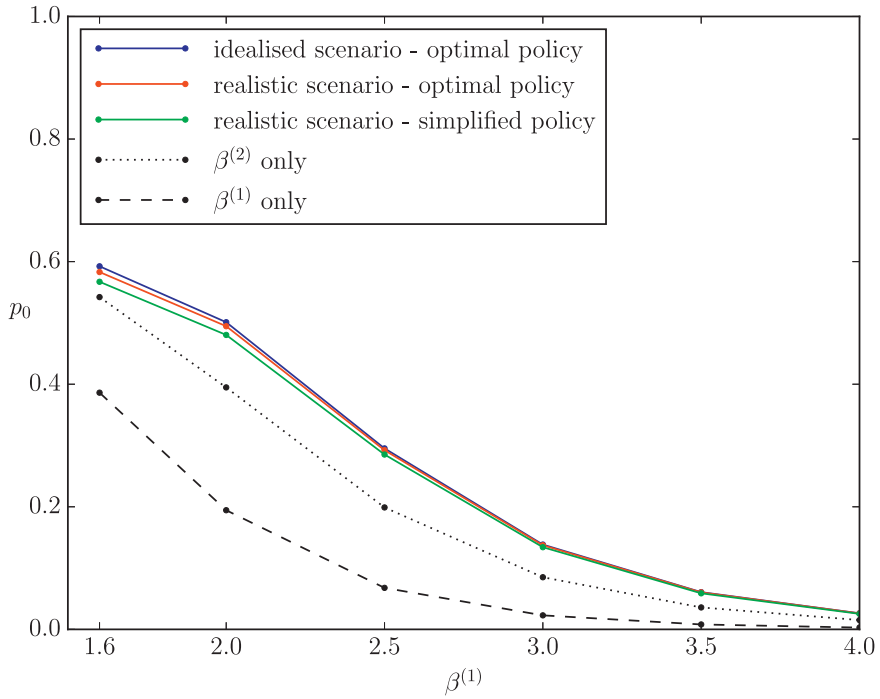


Fig. 5. Comparison of p_0 versus $\beta^{(1)}$ for the three different policies, as well as for no change in β. The parameters are $N = 300$, $\beta^{(2)} = (0.8)\beta^{(1)}$, $\gamma = 1$, $\mu = 0.2$ and $(S_0, I_0) = (N-1, 1)$. $\beta^{(1)} = 3$ corresponds to the policies in Fig. 4. (Lines are in the same vertical order as the legend box.).

5.1. Comparison of optimal policies to the simplified policy

Due to the computational requirements mentioned in Section 3.3, it is only feasible to calculate the optimal policies for small N. Policies were calculated using the methods in Sections 3.2, 4.2 and 3.3.1 for a range of parameters for $N \leq 300$. The corresponding sets V were calculated, and a typical result is shown in Fig. 4. The use of a small N necessitates choosing an unrealistically small value of γ/μ to illustrate the policies. However, the pre-condition (5) is still met.

We see that the idealised and realistic scenarios give very similar policies, and that the simplified policy is a close approximation of both. A similar result was seen with other sets of parameters. These results confirmed the prediction of Section 4.3, that the simplified policy in Section 3.3.1, is a good approximation for either the idealised scenario or the realistic scenario.

The result for the idealised scenario also confirmed the conjecture made in Section 2.2: that the higher β is preferable early in the outbreak, and the lower β is preferable as I falls to a low value. This confirms that in the realistic scenario we should switch from $\beta^{(1)}$ to $\beta^{(2)}$, not the other way around.

For small N it is also possible to calculate p_0 exactly: for the idealised scenario, (12) is solved and then $p_0 = \nu_0(n)$; for the realistic scenario, (27) is solved and then $p_0 = \nu_0$. Some typical results are shown in Fig. 5. It compares the optimal policies for the two different scenarios, as well as the simplified policy under the realistic scenario. Also shown are the outcomes with no change to β.

We see that the results from the three optimisation scenarios are extremely similar. This was a result we observed consistently over a wide range of parameters. This confirms another result of Sections 3.3 and 4.3: that the simplified policy is very nearly as good as the optimal policy, in either scenario.

Therefore we conclude that the simplified policy is a good practical choice, because it is easy to calculate, and so we use it in the further tests in the following sections.

5.2. Effectiveness of the simplified policy

We examine the effectiveness of the simplified policy in the realistic scenario. Note that for large N, exact calculation of p_0 is impractical, so we calculate p_0 using the approximate solution method we previously reported, which has an average error of less than 1% [3].

Fig. 6 shows a typical result, varying $\beta^{(1)}$ and $\beta^{(2)}/\beta^{(1)}$ for a given N, γ and μ. (Although we only show the realistic scenario, the results for the idealised scenario are extremely close, to the point that the plots look identical.) We see that dramatic improvements in p_0 can be achieved with a relatively small reduction in β. Again, we tested a wide range of parameters, and the simplified policy consistently gave significant improvement.

In passing, note that the "$\beta^{(2)}/\beta^{(1)} = 1.0$" curve in Fig. 6 shows non-monotonicity in β, with a local maximum near $\beta/(\gamma+\mu) = 2$, as previously reported [3].

5.3. Comparison of the simplified policy to other simple policies

The simplified policy of (24) is easy to calculate, but a practical problem is that in an outbreak scenario where we might wish to implement our policy, the precise values of the epidemiological parameters are only estimates, and the precise epidemiological status of the population (in terms of the numbers of susceptible and infectious individuals) can once again only be estimated.

Therefore we investigated the robustness of the policy to implementing the intervention at other stages of the epidemic. These results are shown in Fig. 7. We considered the $\beta^{(1)} = 3$ case of Fig. 6,

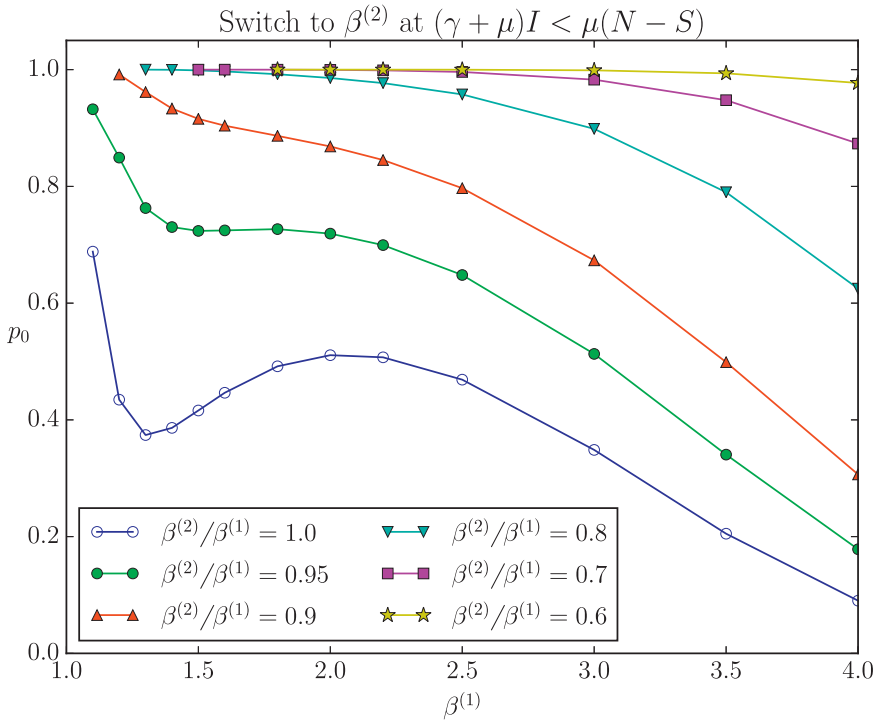


Fig. 6. Plot of p_0 versus $\beta^{(1)}$, for different values of $\beta^{(2)}/\beta^{(1)}$ (realistic scenario, simplified policy). The parameters are $N = 30,000$, $\gamma = 1$, $\mu = 0.025$ and $(S_0, I_0) = (N-1, 1)$.

keeping the same colour and marker scheme, but switched from $\beta = \beta^{(1)}$ to $\beta = \beta^{(2)}$ at other points in the cycle: points *A* through to *E* in Fig. 2. *A* refers to the case of using $\beta^{(2)}$ exclusively; *B* is the deterministic maximum *I* point ($S = S_d^{(1)}$ while $I > I_d^{(1)}$); *C* is when *I* falls to I_e; *D* is when *I* reaches its deterministic local minimum ($S = S_d^{(1)}$ after point *C*) and *E* refers to the case where β is always $\beta^{(1)}$. The simplified policy, denoted by *, is between points *B* and *C*.

As predicted, the simplified policy always gives some improvement over using either $\beta^{(1)}$ or $\beta^{(2)}$ exclusively. However, we found that although the simplified policy gave the best results, points *B* and *C* also gave significant improvement. On the other hand, intervening to reduce the transmission rate parameter, β, too late (point *D*) may or may not be preferable to using $\beta = \beta^{(2)}$ always (point *A*). (In the example in Fig. 7, it is always preferable, but tests with other parameter values have indicated that this is not always the case).

So this shows that there is a wide range of switch points which give some improvement over using either $\beta = \beta^{(1)}$ or $\beta = \beta^{(2)}$ exclusively. So long as a switch point is chosen after the peak *I* (point *B*) and some time before the deterministic local minimum of *I* (point *D*), a significant increase in p_0 will be achievable.

6. Conclusion

In the SIR-with-demography model, reducing the transmission rate parameter from $\beta^{(1)}$ to $\beta^{(2)}$ at an appropriate point can give a substantial increase in the probability of epidemic fade-out, over that when using $\beta^{(1)}$ or $\beta^{(2)}$ exclusively. We believe that this has applications for timing the implementation of epidemic control measures, making it more likely for an epidemic to fade out before it becomes endemic. For instance, if there is a large outbreak, control measures (such as the allocation of antivirals) might be delayed until the epidemic is waning, approximately meeting the condition given by (23).

This method is effective because it allows the epidemic to progress longer without intervention and infect more individuals in the initial outbreak, but this is balanced against the long term gain of epidemic fade-out. As we previously noted, this is not likely to be ethically possible in lethal epidemics. However, it could have applications in situations of non-lethal infections, or diseases among animals, where losses can be economically measured [16]. Possible future research is to investigate the tradeoffs between such policies in terms of total infections. The result of Grigorieva and Khailov with a deterministic SIR model [8], that not reducing β early in the infection cycle can minimise the total infection size, suggests that there might be a similar result when using a stochastic model.

Optimal policies may be calculated using Markov decision process theory, but these are impractical for all but the smallest systems. We have presented a simplified policy (24) which gives a very close to optimal solution. The key factor in determining this policy, the inequality $(\gamma+\mu)I < \mu(N-S)$, is easy to test and is independent of the transmission rate parameter, or the effectiveness of the control measures.

We also observed that even a sub-optimal switch point can give a substantial increase in the probability of epidemic fade-out. This should be useful for practical applications where the exact state of the system is not easily observed.

The method of calculating the simplified policy is based on using the deterministic local minimum to estimate the relative probability of epidemic fade-out. This technique should be amenable to many Markov process problems which concern optimising the probability of hitting one boundary before another. A possibility for future work is to apply this technique to related problems, such as more complicated models, or to the evaluation of the probability of fade-out at other points in the epidemic cycle.

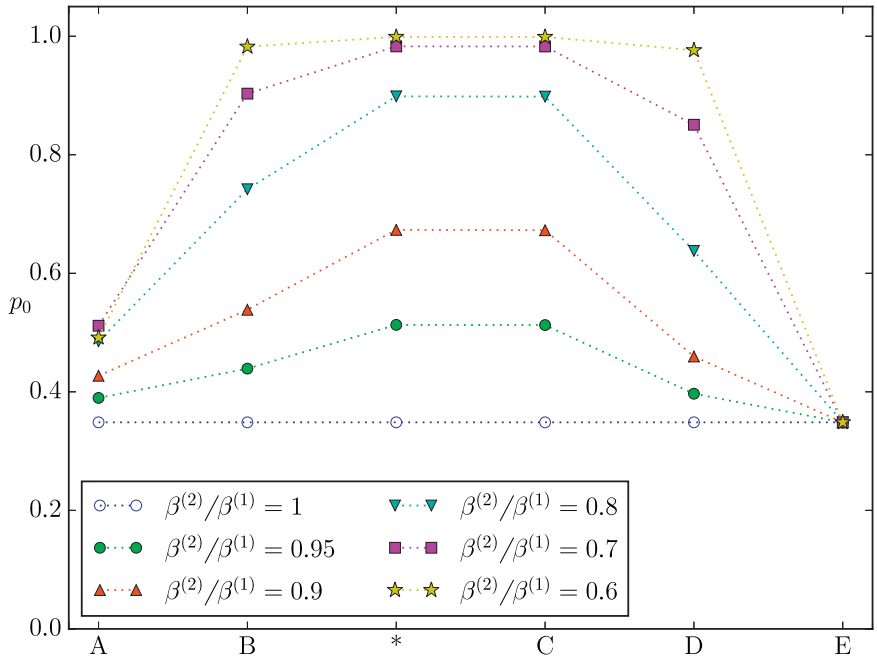


Fig. 7. Comparison of p_0 for different switch points. * is the simplified policy of (24). Points *A* to *E* are shown in Fig. 2. The * results correspond to the $\beta^{(1)} = 3$ results in Fig. 6. The parameters are $N = 30,000$, $\beta^{(1)} = 3$, $\gamma = 1$, $\mu = 0.025$ and $(S_0, I_0) = (N-1, 1)$.

Acknowledgments

This work is supported by an APA Scholarship (PB), an Australian Research Council Future Fellowship (JVR; FT130100254), and the NHMRC (JVR; CRE PRISM^2).

References

[1] R.M. Anderson, R.M. May, Infectious Diseases of Humans: dynamics and control, Oxford University Press, Oxford, 1991.
[2] W.J. Anderson, Continuous-time Markov chains : an applications-oriented approach, Springer-Verlag, New York, 1991.
[3] P.G. Ballard, N.G. Bean, J.V. Ross, The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography, J. Theor Biol 393 (2016) 170–178, doi:10.1016/j.jtbi.2016.01.012.
[4] A.J. Black, T. House, M.J. Keeling, J.V. Ross, Epidemiological consequences of household-based antiviral prophylaxis for pandemic influenza, J. R. Soc. Interface 10 (2013) 20121019, doi:10.1098/rsif.2012.1019.
[5] T. Britton, T. House, A.L. Lloyd, D. Mollison, S. Riley, P. Trapman, Five challenges for stochastic epidemic models involving global transmission, Epidemics 10 (2015) 54–57, doi:10.1016/j.epidem.2014.05.002.
[6] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, John Wiley & Sons, Chichester, 2000.
[7] Z. Feng, S. Towers, Y. Yang, Modeling the effects of vaccination and treatment on pandemic influenza, AAPS J. 13 (3) (2011) 427–437, doi:10.1208/s12248-011-9284-7.
[8] E.V. Grigorieva, E.N. Khailov, Optimal vaccination, treatment, and preventive campaigns in regard to the SIR epidemic model, Math. Model. Nat. Phenom 9 (4) (2014) 105–121, doi:10.1051/mmnp/20149407.
[9] A. Handel, I.M. Longini, R. Antia, Antiviral resistance and the control of pandemic influenza: The roles of stochasticity, evolution and model details, J. Theor. Biol. 256 (2009) 117–125, doi:10.1016/j.jtbi.2008.09.021.
[10] C. Jackson, P. Mangtani, J. Hawker, B. Olowokure, E. Vynnycky, The effects of school closures on influenza outbreaks and pandemics: systematic review of simulation studies, PLoS ONE 9 (5) (2014) e97297, doi:10.1371/journal.pone.0097297.
[11] A. Kamenev, B. Meerson, Extinction of an infectious disease: A large fluctuation in a nonequilibrium system, Phys. Rev. E. 77 (2008) 061107, doi:10.1103/PhysRevE.77.061107.
[12] T.K. Kar, A. Batabyal, Stability analysis and optimal control of an SIR epidemic model with vaccination, BioSystems 104 (2011) 127–135, doi:10.1016/j.biosystems.2011.02.001.
[13] E. Kreyszig, Advanced Engineering Mathematics, 8th, John Wiley & Sons, New York, 1999.
[14] T.G. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, J. Appl. Probab. 7 (1) (1970) 49–58, doi:10.2307/3212147.
[15] S. Lee, G. Chowell, Exploring optimal control strategies in seasonally varying flu-like epidemics, J. Theor. Biol. (2016). http://dx.doi.org/10.1016/j.jtbi.2016.09.023.
[16] C. Lefèvre, Optimal control of a birth and death epidemic process, Oper. Res. 29 (5) (1981) 971–982.
[17] I.M. Longini, M.E. Halloran, A. Nizam, Y. Yang, Containing pandemic influenza with antiviral agents, Am. J. Epidemiol 159 (7) (2004) 623–633, doi:10.1093/aje/kwh092.
[18] M. Lydeamore, N.G. Bean, A.J. Black, J.V. Ross, Choice of antiviral allocation scheme for pandemic influenza depends on strain transmissibility, delivery delay and stockpile size, Bull. Math. Biol. 78 (2) (2016) 293–321, doi:10.1007/s11538-016-0144-6.
[19] B. Meerson, P.V. Sasorov, WKB theory of epidemic fade-out in stochastic populations, Phys. Rev. E 80 (2009) 041130, doi:10.1103/PhysRevE.80.041130.
[20] I. Nåsell, Stochastic models of some endemic infections, Math. Biosci. 179 (2002) 1–19, doi:10.1016/S0025-5564(02)00098-6.
[21] R.L.M. Neilan, E. Schaefer, H. Gaff, K.R. Fister, S. Lenhart, Modeling optimal intervention strategies for cholera, Bull. Math. Biol. 72 (2010) 2004–2018, doi:10.1007/s11538-010-9521-8.
[22] A.B. Piunovskiy, D. Clancy, An explicit optimal intervention policy for a deterministic epidemic model, Optim. Control Appl. Methods 29 (2008) 413–428, doi:10.1002/oca.834.
[23] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley, New York, 1994.
[24] G. Rozhnova, C.J.E. Metcalf, B.T. Grenfell, Characterizing the dynamics of rubella relative to measles: the role of stochasticity, J. R. Soc. Interface 10 (2013) 20130643.
[25] M.W. Tanner, L. Sattenspiel, L. Ntaimo, Finding optimal vaccination strategies under parameter uncertainty using stochastic programming, Math. Biosci. 215 (2008) 144–151, doi:10.1016/j.mbs.2008.07.006.
[26] O.A. van Herwaarden, Stochastic epidemics: the probability of extinction of an infectious disease at the end of a major outbreak, J. Math. Biol. 35 (1997) 793–813.
[27] Y. Xiao, F. Brauer, S. Moghadas, Can treatment increase the epidemic size? J. Math. Biol. 72 (2016) 343–361, doi:10.1007/s00285-015-0887-y.
[28] R. Yaesoubi, T. Cohen, Generalized Markov models of infectious disease spread: a novel framework for developing dynamic health policies, Eur. J. Oper. Res. 215 (2011) 679–687, doi:10.1016/j.ejor.2011.07.016.

5 Paper 3

5.1 Introduction

Paper 3 is entitled "The impact of time dependent transmission rate on the probability of epidemic fade-out" [8]. It was submitted to *Journal of Theoretical Biology* in early 2018, and is currently (March 2018) under review.

This paper extended the work of Paper 1 by calculating p_0 when the transmission rate parameter, β, is time dependent. This is a natural extension which had been suggested both by ourselves [6] and by others [19]. In addition to the SIR-with-demography model, which was the sole model in the first two papers, we extended the method to also account for a slightly different model, the SIRS model,

Finally, we applied this method by examining trends in p_0 for two realistic parameter sets: influenza-like parameters and measles-like parameters.

5.2 Statement of Authorship

Statement of Authorship

Title of Paper	The impact of time dependent transmission rate on the probability of epidemic fade-out
Publication Status	**Submitted for publication**
Publication Details	Submitted to Journal of Theoretical Biology, February 2018

Principal Author

Name of Principal Author (Candidate)	Peter Ballard		
Contribution to the Paper	Derived all the mathematics, wrote all the algorithms and code, ran all the code, generated all the diagrams, wrote most of the text		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	28-3-2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Prof. Nigel Bean		
Contribution to the Paper	Project supervision, idea generation, suggestions and corrections to text.		
Overall percentage (%)	10%		
Signature		Date	28/03/2018

Name of Co-Author	Prof. Joshua Ross		
Contribution to the Paper	Project supervision, idea generation, helped choose parameters, suggestions and corrections to text.		
Overall percentage (%)	10%		
Signature		Date	28/03/18

5.3 Paper 3

The paper, as submitted to *Journal of Theoretical Biology* in 2018, is on the following pages.

The impact of time-dependent transmission rate on the probability of epidemic fade-out

P. G. Ballard[1,*], N. G. Bean[1,], J. V. Ross[1,]

Abstract

Epidemic fade-out refers to an infection fading out in the first trough after the initial wave of a major outbreak. Previous work on the probability of epidemic fade-out has used models with constant parameters. We present a technique for efficiently calculating the probability of epidemic fade-out with a time-dependent transmission rate parameter. A general analysis reveals the different effects on this probability. We then apply the method to two different scenarios: we consider influenza-like parameters with a seasonal variation in transmission rate; and we consider measles-like parameters with the transmission rate changing during school vacations. Both scenarios show that the probability of epidemic fade-out depends on the time of year at which the epidemic is introduced, but in different ways. For flu-like parameters we observe that the probability of epidemic fade-out has a local minimum at a higher R_0 value than the previously reported $R_0 \approx 2$ using constant parameters. For measles-like parameters, the seasonal variation lowers the average probability of epidemic fade-out.

1. Introduction

In modelling the outbreak of an infectious disease, a relatively rarely studied effect is the phenomenon of *epidemic fade-out*. This refers to the case when an infection fades out after a large initial outbreak [1]. This is in contrast to *initial fade-out*, when it fades out before a major outbreak occurs [9], and *endemic fade-out*, where the infection survives the initial stages, reaches an endemic state, and fades out at some later time [1].

Epidemic fade-out has been suggested as an area which is deserving of further research [13, 8, 9]. We are aware of only two papers which give approximate formulae for the probability of epidemic fade-out [32, 24], as well as our own numerical approximation [3]; and our study into maximising this probability [4]. However, all of these works assume constant parameters.

*Corresponding author; peter.ballard@adelaide.edu.au

[1]School of Mathematical Sciences, and ARC Centre of Excellence for Mathematical and Statistical Frontiers, The University of Adelaide, Adelaide SA 5005, AUSTRALIA.

In any realistic system, some of the parameters will change through time. In particular it has been suggested [9] that a useful step, in analysing epidemic fade-out, would be to account for the fact that the transmission rate parameter can vary with time.

Therefore we present here the first study into epidemic fade-out without assuming time-invariant parameters. We present a numerical computation method for calculating the probability of epidemic fade-out, which accounts for a time-dependent transmission rate parameter. Obviously this calculation could also be done with Monte Carlo simulations, but the required number of simulations is very large; and the numerical method allows computations to be done relatively quickly and in a systematic way. In this paper, Monte Carlo simulations are still done as "spot checks" on a small number of the results, to verify the accuracy of the method.

We then perform a general analysis of the effect of time dependence at different times on the outbreak cycle. Finally, we apply it to two particular examples.

The first example is for influenza-like parameters. Although influenza generally is seasonal and fades out in a community near the end of winter in temperate climates [33], there are exceptions. It has been observed in Australia that influenza outbreaks have a more complex pattern in the tropical regions [31]. Furthermore, some severe influenza outbreaks have been observed to occur in multiple waves in some regions, rather than a single outbreak. This was observed both for the 1918-1919 Spanish flu pandemic [25], and for the 2009 swine flu outbreak in the northern hemisphere [14, 30]. Therefore, changes in the transmission rate parameter could prove important in understanding the dynamics of an outbreak of influenza or a similar disease.

The second example is one of an illness which usually carries lifetime immunity, such as measles. Modelling seasonal variation in measles is not new [5, 18], however we believe our approach is novel in that we use a stochastic model to survey trends in epidemic fade-out. We use a model where the transmission rate parameter changes due to school vacations, which appear to be the primary driver of variation for measles [16]; though it may be that seasonal factors are more important in less-developed countries [10].

Both examples show strong dependence on the time of year at which an outbreak is introduced. For the influenza example, we also see dependence on R_0, the basic reproduction number. Previously [3] we had reported that the probability of surviving epidemic fade-out has a local minimum near $R_0 = 2$. We note that, with seasonal forcing, this local minimum occurs at a higher value of R_0, dependent on the amplitude of the seasonal forcing. For measles-like parameters, we see that the imposition of seasonal forcing moves the simulated Critical Community Size (CCS) closer to the observed value [6] for measles.

2. Models and definitions

We use the Markovian SIR-with-demography model [27], and the Markovian SIRS (Susceptible-Infectious-Recovered-Susceptible) model [20]. These both extend the well-known Markovian SIR infection model by allowing replenishment

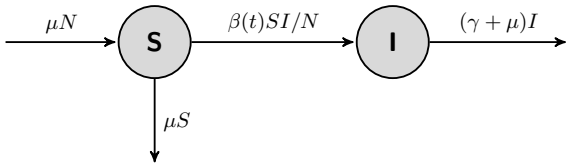


Figure 1: The SIR-with-demography epidemic model. S is the number of susceptibles and I is the number of infectious individuals. $\beta(t)$ is the time-dependent transmission rate parameter. The other parameters are fixed: the population size parameter N, the recovery rate γ, and the per-capita birth/death rate μ.

of susceptibles, though in different ways. These models are simple, not accounting for factors such as latent periods or heterogeneous populations, but their simplicity allows trends to be more easily analysed.

The SIR-with-demography model replenishes susceptibles via births and deaths, or immigration and emigration. It assumes individuals have lifetime immunity, so is appropriate for modelling diseases such as measles. The SIRS model replenishes susceptibles by the waning of immunity, and is more suitable for modelling disease outbreaks where the duration of immunity is short compared to individuals' lifetime, as is the case for influenza.

The models are closely related and have very similar characteristics. The SIR-with-demography model tends to be slightly more complicated because the population size is not fixed and there is one more type of event. So we shall mainly present theory for the SIR-with-demography model, then note the differences for the SIRS model.

2.1. The SIR-with-demography model

The Continuous-Time Markov Chain (CTMC) for the SIR-with-demography model is described in Figure 1 and Table 1. S and I represent the number of "susceptible" and "infectious" individuals respectively. The number of "recovered" individuals (R) does not affect S and I, so it can be removed from the analysis [19]. This allows "death of infectious" (at rate μI) and "recovery of infectious" (at rate γI) to be grouped together as "removal of infectious".

Description	Transition	Rate
Infection	$(S, I) \to (S-1, I+1)$	$\beta(t)SI/N$
Birth of Susceptible	$(S, I) \to (S+1, I)$	μN
Death of Susceptible	$(S, I) \to (S-1, I)$	μS
Removal of Infectious	$(S, I) \to (S, I-1)$	$(\gamma+\mu)I$

Table 1: Transition rates for the Markovian SIR-with-demography epidemic model displayed in Figure 1.

We consider cases where the infection rate parameter β can vary with time, so we denote it as a function of time, $\beta(t)$.

2.1.1. Preliminary theory

In the limit as N becomes large, the stochastic process converges to a deterministic process [29], which may also be called the deterministic approximation of the stochastic process [26]. It is governed by the differential equations:

$$\frac{dS}{dt} = \mu(N - S) - \frac{\beta(t)SI}{N}, \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta(t)SI}{N} - (\gamma + \mu)I. \tag{2}$$

At any point in time, the attracting fixed point (at which both (1) and (2) are zero) is the point $(S_e(t), I_e(t))$, where:

$$S_e(t) = \frac{N(\gamma + \mu)}{\beta(t)}, \tag{3}$$

$$I_e(t) = \frac{N\mu}{\gamma + \mu}\left(1 - \frac{\gamma + \mu}{\beta(t)}\right). \tag{4}$$

When $\beta(t)$ is constant, the meaning of the attracting fixed point is obvious: it corresponds to the endemic point, which the deterministic process converges towards in an anticlockwise direction in the (S, I) plane, as illustrated in Figure 2. Again referring to Figure 2, a stochastic realisation may fade out at the start, near point A (initial fade-out), or in the first trough after the initial outbreak, near point E (epidemic fade-out).

When $\beta(t)$ is variable, it makes less sense to speak of an endemic point. But the position of $(S_e(t), I_e(t))$ still governs the behaviour of the deterministic process. At any point in time, $(S_e(t), I_e(t))$ is an attracting point, causing the deterministic process at that point in time to move towards it in an anticlockwise direction.

The stochastic model has discrete states, so it is also useful to round the endemic state values up to the next highest integer pair:

$$S_d(t) = \lceil S_e(t) \rceil, \tag{5}$$

$$I_d(t) = \lceil I_e(t) \rceil. \tag{6}$$

In a naive population, $S \approx N$. So assuming a constant β, (2) gives R_0, the basic reproduction number, to be:

$$R_0 = \frac{\beta}{\gamma + \mu}. \tag{7}$$

We impose the condition $R_0 > 1$, because that is a requirement for a major outbreak to occur.

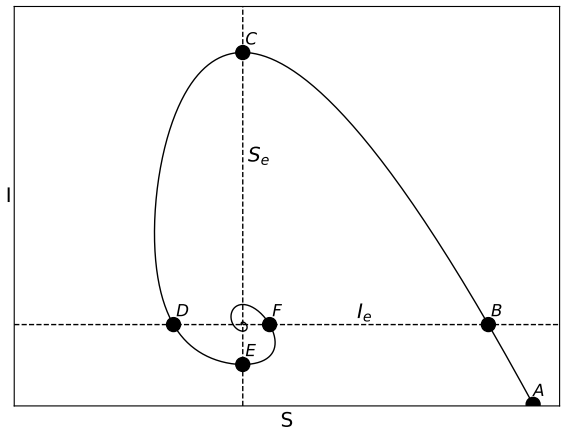


Figure 2: I versus S plot of the deterministic approximation of a typical SIR-with-demography or SIRS model, with endemic values S_e and I_e (dashed lines). The outbreak starts at A, goes through points B, C, D and E through to F, and converges on (S_e, I_e). The *first trough* is between points D and F.

A stability analysis of (1) and (2), linearised in the region of (S_e, I_e) as given by (3) and (4), determines that the associated eigenvalues λ obey,

$$\lambda^2 + R_0\mu\lambda + \mu(\gamma+\mu)(R_0-1) = 0.$$

This gives the eigenvalues,

$$\lambda = \frac{-R_0\mu}{2} \pm \left(\frac{\mu}{2}\right)\sqrt{{R_0}^2 - 4(R_0-1)(\gamma+\mu)/\mu}.$$

In almost any realistic case (and in all cases considered in this paper), the recovery rate γ is much greater than the population turnover rate μ. This means $\gamma \gg \mu$, and hence $(R_0-1)(\gamma+\mu)/\mu \gg {R_0}^2$, so this simplifies to,

$$\lambda \approx \frac{-R_0\mu}{2} \pm i\sqrt{(R_0-1)(\gamma+\mu)\mu}.$$

In general, eigenvalues of the form $-a \pm ib$ correspond to a decay time constant of $1/a$ and oscillations with a period of $2\pi/b$ [21]. So in this case, in the limit as the process approaches the endemic point, the deterministic trajectory spirals in to the endemic point, circling with period,

$$T_{osc} \approx \frac{2\pi}{\sqrt{(R_0-1)(\gamma+\mu)\mu}}. \tag{8}$$

Even though we are considering a stochastic system, with time-varying β (and hence time-varying R_0), (8) is a useful first-order approximation for the period of oscillations, which we shall use in Section 4.

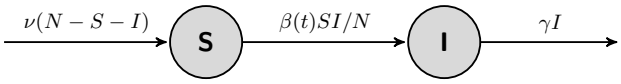


Figure 3: The SIRS model. S is the number of susceptibles and I is the number of infectious individuals. $\beta(t)$ is the time-dependent transmission rate parameter. The other parameters are fixed: the population size N, the recovery rate γ, and the immunity waning rate ν.

2.2. The SIRS model

Description	Transition	Rate
Infection	$(S, I) \to (S-1, I+1)$	$\beta(t)SI/N$
Removal of Infectious	$(S, I) \to (S, I-1)$	γI
Loss of Immunity	$(S, I) \to (S+1, I)$	$\nu(N-S-I)$

Table 2: Transition rates for the Markovian SIRS epidemic model displayed in Figure 3.

The SIRS (Susceptible-Infectious-Recovered-Susceptible) model [20] is described in Figure 3 and Table 2. Recovered individuals regain susceptibility at per capita rate ν, corresponding to the waning of their immunity. The population size is fixed at N, so the number in the "recovered" class is equal to $N-S-I$, giving a replenishment rate of $\nu(N-S-I)$.

The analysis for the SIR-with-demography model in Section 2.1.1 all holds, with the exception that Equations (1), (2), (3), (4), (7) and (8) are instead:

$$\frac{dS}{dt} = \nu(N-S-I) - \frac{\beta(t)SI}{N}, \tag{9}$$

$$\frac{dI}{dt} = \frac{\beta(t)SI}{N} - \gamma I, \tag{10}$$

$$S_e(t) = \frac{N\gamma}{\beta(t)}, \tag{11}$$

$$I_e(t) = \frac{N\nu}{\gamma+\nu}\left(1 - \frac{\gamma}{\beta(t)}\right), \tag{12}$$

$$R_0 = \frac{\beta}{\gamma}, \tag{13}$$

$$T_{osc} \approx \frac{2\pi}{\sqrt{(R_0-1)\gamma\nu}}. \tag{14}$$

2.3. Definition of epidemic fade-out

While we may say informally that epidemic fade-out refers to fade-out during the first trough after the initial wave of a major outbreak, a precise definition is needed for calculations of its probability.

In a previous paper [3], we defined the first trough as starting when the CTMC enters a state for which $S < S_d$ and $I = I_d$ (near point D in Figure 2). Then either absorption occurs (and the infection fades out) at the lower absorbing boundary L (specified by $I = 0$), or the CTMC exits the trough when it reaches an artificial upper absorbing boundary U, which normally occurs when $S \geq S_d$ and $I \geq I_d$ (near point F in Figure 2).

For this work, extra specifications are required, because we need to account for the fact that $S_d(t)$ and $I_d(t)$ can change, since $\beta(t)$ is not constant. Therefore, if $\mathbb{S}$ is the state space of all possible (S, I) values, we specify a first trough region T, a lower absorbing boundary L, and an artificial upper absorbing boundary U, in terms of the *variable* values $S_d(t)$ and $I_d(t)$:

$$T = \{(S, I) \in \mathbb{S} | I < I_d(t)\}; \tag{15}$$

$$L = \{(S, I) \in \mathbb{S} | I = 0\};$$

$$U = \{(S, I) \in \mathbb{S} | (S \geq S_d(t) \text{ and } I \geq I_d(t))\}.$$

Note that T and U are calculated instantaneously. That is, the first trough is entered at the first point at which (15) is true, given the $\beta(t)$ in use at that instant. Similarly, absorption at L or U is calculated instantaneously (although L never changes). The justification is that, at any instant in time, the CTMC is converging towards the endemic point specified by its current parameters.

In addition, we need to avoid a situation where an outbreak is defined to have occurred simply through a change in $S_d(t)$. This may occur in situations where the trajectory converges directly towards the endemic point rather than oscillate around it; meaning one can not meaningfully speak of epidemic fade-out. Therefore we introduce and use the set Q to determine whether a major outbreak has occurred. That is, we let

$$S_{dmin} = \min_{t \geq 0} S_d(t),$$

and

$$Q = \{(S, I) \in \mathbb{S} | S < S_{dmin}\}; \tag{16}$$

a pre-condition of a major outbreak occurring is that the CTMC has entered Q.

Regions Q, T, L and U are shown in Figure 4. We then specify p_0, the conditional probability of epidemic fade-out, as the probability that the process is absorbed at L before reaching a state in U, given that it reaches Q and then T. We also specify p_1, the probability of surviving epidemic fade-out, to be $1 - p_0$.

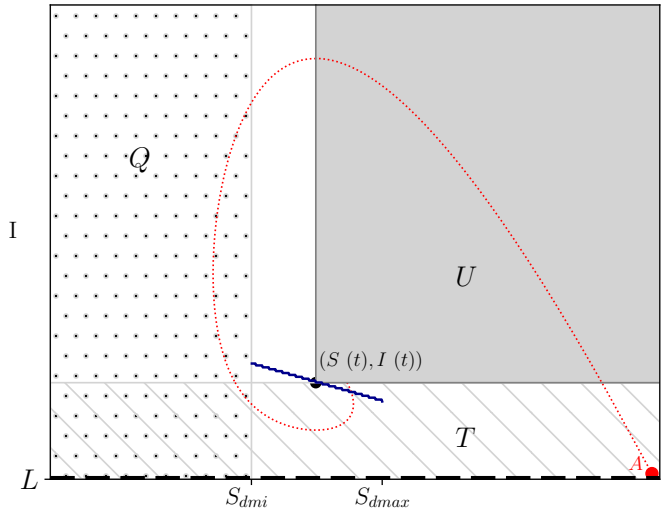


Figure 4: Graphical representation of the definition of epidemic fade-out: the condition is that the CTMC must enter Q ($S < S_{dmin}$, the dotted region) and then T ($I < I_d(t)$, the hatched region). Given that condition, the probability of epidemic fade-out (p_0) is the probability of the CTMC being absorbed at L ($I = 0$, the thick dashed line) before next reaching U ($S \geq S_d(t)$ and $I \geq I_d(t)$, the grey region). Both T and U may change, with the vertex $(S_d(t), I_d(t))$ able to move along the dark blue solid line as $\beta(t)$ varies. The red (dotted) line shows a possible behaviour of the deterministic approximation, starting at point A.

3. Method of calculating the probability of epidemic fade-out

In Section 3.3.2 of a previous paper [3], we described a fast and accurate numerical method for calculating the probability of epidemic fade-out in the SIR-with-demography infection model, with constant parameters. Here we describe a modification of this algorithm, which allows time-dependent parameters.

We break the algorithm into two parts: a continuous diffusion approximation when I is large (Section 3.1), and a semi-continuous approximation (discrete I but continuous S) when I is small (Section 3.2). We first describe these two methods, without specifying what constitutes "large" or "small" I, before considering how to best combine them in Section 3.3.

3.1. Diffusion approximation

The diffusion approximation is an extension of the deterministic approximation. It tracks both the mean and the variance of the state with time, while the deterministic approximation tracks the mean state only. Aside from the hitting time variance (17), this is outlined in more detail in [3].

Let $Y_N(t)$ be a process following the SIR-with-demography model defined in Section 2.1 (calculations for the SIRS model in Section 2.2 are analogous), with each value being an (S, I) pair, from some initial value (S_0, I_0). The associated *density process* is $X_N(t) = Y_N(t)/N$, with each possible value x being an (s, i) pair, where $s = S/N$ and $i = I/N$; and the initial value is $x_0 = (s_0, i_0) = (S_0/N, I_0/N)$.

Let $f(x, l)$ be the transition rate of the density process from state (x) to state $(x + l/N)$, where $l = (l_1, l_2)$ can take on the possible 1-step transition values in Table 1: $(-1, 1)$, $(1, 0)$, $(-1, 0)$ and $(0, -1)$, respectively. Also define for the density

$$F(x) = \sum_l l f(x, l) = (-\beta(t)si + \mu(1 - s), \quad \beta(t)si - (\gamma + \mu)i)\,;$$

$B(x)$, a matrix whose $(j, k)^{th}$ element is given by $b_{j,k} = \frac{\partial f_j}{\partial x_k}$,

$$\Rightarrow B(x) = \begin{pmatrix} -\beta(t)i - \mu & -\beta(t)s \\ \beta(t)i & \beta(t)s - (\gamma + \mu) \end{pmatrix};$$

and $G(x)$, a matrix whose $(j, k)^{th}$ element is given by $g_{j,k} = \sum_l \; l_j \; l_k \; f(x, l)$,

$$\Rightarrow G(x) = \begin{pmatrix} \beta(t)si + \mu(1 + s) & -\beta(t)si \\ -\beta(t)si & \beta(t)si + (\gamma + \mu)i \end{pmatrix}.$$

Then by Theorem 3.1 of Kurtz [22] and Theorem 3.2 of Pollett [28], we have: *In the limit as $N \to \infty$, $X_N(t)$ converges weakly (on $D[0, T]$, the space of right-continuous, left-hand limits functions on $[0, T]$) to a process which at time t is Gaussian with mean $X(t)$ and covariance $\Sigma(t)/N$; where $X(t)$ and $\Sigma(t)$ are the solutions to:*

$$\frac{dX(t)}{dt} = F(X(t)), \quad X(0) = (s_0, i_0);$$

$$\frac{d\Sigma(t)}{dt} = B(X(t))\Sigma(t) + \Sigma(t)B(X(t))^T + G(X(t)), \quad \Sigma(0) = \mathbf{0}.$$

If $Y_N(t) = NX_N(t)$ reaches the T defined in (15) at time τ, i.e. $\tau = \min\{t \geq 0 : Y_N(t) \in T\}$, then by applying Theorem 11.4.1 of Ethier and Kurtz [15] we have: *In the limit as $N \to \infty$, the distribution of the scaled density process $Y_N(t)$ when it first enters T is Gaussian; with S having mean $NX(\tau)_1$ and variance*

$$\sigma_S^2 = N\left(\Sigma(\tau)_{1,1} + \left(\frac{F(X(\tau))_1}{F(X(\tau))_2}\right)^2 \Sigma(\tau)_{2,2} - 2\left(\frac{F(X(\tau))_1}{F(X(\tau))_2}\right)\Sigma(\tau)_{1,2}\right) \quad ;$$

and with hitting time having mean τ, and variance

$$\sigma_\tau^2 = N\left(\frac{\Sigma(\tau)_{2,2}}{F(X(\tau))_2}\right); \tag{17}$$

(where subscript j denotes the j^{th} element of a vector, and subscript j,k denotes the row j, column k element of a matrix).

3.2. Semi-continuous approximation

The diffusion approximation does not take account of absorbing boundaries. So when I is low, meaning the CTMC is near the $I = 0$ absorbing boundary, a more accurate technique is required. However the state space is too large to perform an exact Markov calculation. What we desire is an accurate approximate method.

The approximation we use is to treat S as continuous, but keep I discrete. The justification for this is that I is the critical dimension: absorption at L always occurs in the I dimension, at $I = 0$; and absorption at U almost always occurs in the I dimension, at $S > S_d$ and $I = I_d$. Therefore the I dimension is more critical, and treating S as continuous should not come at a great cost in accuracy. This allows an iterative process in which, on each step, a time step-size is chosen, and S is incremented by dS/dt multiplied by that time step. This includes the assumption that, at any time, all probability mass shares the same S value. This is obviously a significant assumption, but it has the advantage that it simplifies the calculation, because only a one-dimensional vector of probability mass values needs to be stored at any one time.

We define I_u to be the instantaneous lowest state in the upper absorbing boundary U; and we define I_{max} to be the maximum allowable value of I_u.

We define a vector, E, of probability mass. That is, element i of E, denoted E_i, contains the probability of $I = i$. E contains elements 0 through to I_{max}, with elements 0 and I_{max} representing the absorbing states.

With a deterministic model for S, but a stochastic model for I, we need a "best" I to use when calculating dS/dt. We use I_{mean}, the expected I conditioned on non-absorption:

$$I_{mean} = \frac{\sum_{I=1}^{I_u-1} IE_I}{\sum_{I=1}^{I_u-1} E_I}.$$

So the differential equation for the common S is, for the SIR-with-demography model:

$$\frac{dS}{dt} = \mu(N - S) - \beta S I_{mean}/N; \tag{18}$$

and for the SIRS model:

$$\frac{dS}{dt} = \nu(N - S - I_{mean}) - \beta S I_{mean}/N. \tag{19}$$

Let us define $\Gamma = \gamma + \mu$ for the SIR-with-demography model, and $\Gamma = \gamma$ for the SIRS model. Then for each state E_I with I in the range $0 < I < I_u$, there is a process moving probability mass into state E_{I+1} at rate $\beta SI/N$, and a process moving probability mass into state E_{I-1} at rate ΓI.

With this, we can now specify differential equations for E:

$$\frac{dE_I}{dt} = \begin{cases} (\beta S/N) I_u E_{I_u} & \text{if } I = I_{max}, \\ (\beta S(I-1)/N) E_{I-1} - (\beta SI/N) E_I - \Gamma I E_I & \text{if } I = I_u - 1, \\ (\beta S(I-1)/N) E_{I-1} + \Gamma(I+1) E_{I+1} - (\beta SI/N) E_I - \Gamma I E_I & \text{if } 0 < I < I_u - 1, \\ \Gamma(I+1) E_{I+1} & \text{if } I = 0. \end{cases} \tag{20}$$

For states with I in the range $I_u \leq I < I_{max}$, all probability mass is moved instantly into state $E_{I_{max}}$.

Differential equations (18) or (19), and (20), can then be evaluated numerically, with a DE solver running until nearly all of the probability mass is in one of the two absorbing states (E_{Imax} and E_0). In the computations in Section 4, we stopped evaluation when more than $1 - 10^{-9}$ of probability mass was in the absorbing states.

The distribution of S and time at either absorbing boundary can also be stored, and passed to the next stage of the calculation.

3.3. Putting it together

3.3.1. Sampling

The algorithm uses the semi-continuous method for the initial stage of the outbreak, when I is low; the diffusion method in the second stage, when I is high; and the semi-continuous method again in the third and final stage, in the trough after the initial outbreak, when I is low again.

However a complication is that the methods do not "interface" simply. The diffusion approximation requires an initial point mass or Gaussian distribution, but it follows the semi-continuous approximation of the initial stage, whose output distribution is non-symmetric. Then the final semi-continuous approximation begins with a point mass in S and time, but it follows the diffusion approximation whose output is a Gaussian distribution.

In both cases, we solve this problem by taking multiple samples of the previous stage. So for the diffusion, the output of the first stage is divided up in time into a number of equal quantiles, and for each quantile, the mean S and time are used for the starting point of the diffusion stage. Ten quantiles were used,

which is a relatively large number, but the diffusion stage has a fast execution time compared to the semi-continuous stages, so the cost is low.

The results of the multiple runs of the diffusion are mixed, giving a probability mass distribution in both S and time. This distribution is then assumed to be Gaussian. (This need not be true in all cases – the output of each diffusion run is Gaussian, but their mixture need not be – but we observed they were consistently close to Gaussian, as well as having only a small covariance between S and time).

We then take samples to approximate this distribution. Even 4 samples are sufficient, so to preserve mean and variance we sample (S, t) at $(\mu_s \pm \sigma_s, \mu_t \pm \sigma_t)$, where μ_s and σ_s are the mean and standard deviation of S, and μ_t and σ_t are the mean and standard deviation of time. The final semi-continuous stage is then run for these 4 samples, and the results are mixed to give a final estimate for p_0.

3.3.2. Boundaries

The final decision is of the boundaries to use between the three stages, and the values of I_u and I_{max} to use in the semi-continuous stages.

First let us define the maximum possible value of $I_d(t)$,

$$I_{dmax} = \max_{t \geq 0} I_d(t).$$

The end points of the first and second stages are not critical, but complications are avoided if a constant value is used for both. For convenience we use the same value: we set $I_u = I_{max} = I_{dmax}$ during the first stage, and end the second (diffusion) stage when I falls to $I = I_{dmax}$.

The third stage is a little more complicated. As discussed in Section 2.3, the absorbing boundary needs to be the time-dependent $I_d(t)$ when $S \geq S_d(t)$. Referring to Figure 4, in the region where $S < S_d(t)$, I is unbounded for the SIR-with-demography model, and can go as high as $N - S$ in the SIRS model; suggesting a large value of I_{max} in either case. But since in that region the recovery rate is greater than the infection rate (that is, the deterministic approximation decreases in I), in practice we only require I_{max} to be large enough so that there is some buffer to account for stochastic variation in I when $I < I_d(t)$. The following was found to be sufficient:

$$I_{max} = \max\left(2I_{dmax}, I_{dmax} + 100\right);$$

$$I_u(t) = \begin{cases} I_{max} & \text{if } S < S_d(t), \\ I_d(t) & \text{if } S \geq S_d(t). \end{cases}$$

4. Results

We now use this method to examine the effect of varying β on the probability of epidemic fade-out.

We note that the initial stage calculates the probability of initial fade-out, as a by-product. These results are not presented here because they are not of direct interest to us; but we note that some checks revealed that this was also consistently very accurate; and that the previous work of Ball [2] was also very accurate.

Note that all figures in this section plot p_1, where $p_1 = 1 - p_0$, to allow for easier presentation.

4.1. Performance

The method was run on a 3 GHz Intel i5 core machine, 8 GB RAM, with Mac OS 10.12.5. The code was implemented in Cython, allowing C-like performance on the critical loops [7]. Typical calculation times are shown in Table 3, These are compared to the times for $10,000$ Monte Carlo simulations (Gillespie method [17]). This number was chosen because it gives $2\sigma \approx 0.01$ for $p_1 \approx 0.5$, so it gives a comparable level of accuracy in the results; though a higher number of simulations is required for p_1 very close to 0 or 1, as in some of the traces in Figures 6 to 9. Table 3 shows that the algorithm tends to be about 20 to 40 times faster than Monte Carlo simulations.

Section	N	Section 3 algorithm	Monte Carlo
4.2, 4.3	20,000	8	360
4.4	200,000	9	250
4.4	500,000	23	600
4.4	1,000,000	60	1300
4.4	2,000,000	180	2700

Table 3: Typical calculation times, in seconds, for the values of N used in Sections 4.2, 4.3 and 4.4.

4.2. Single Pulse

As a first step, we examine the effect of modulation (increase in the value of β) in a single short time span, to build a picture of the effect of modulation at different times in the infection cycle.

To do this, we modulate for a time of $T_{osc}/100$. That is, for start-of-modulation time τ, and some transmission rate parameter β_0,

$$\beta = \begin{cases} 1.1\beta_0 & \text{if } \tau \leq t < \tau + T_{osc}/100; \\ \beta_0 & \text{otherwise.} \end{cases}$$

The results are broadly the same across all parameters, with the main difference being the time scale (due to different values of T_{osc} in (8) or (14)). So we show in Figure 5 a representative plot, using numbers similar to Section 4.3:

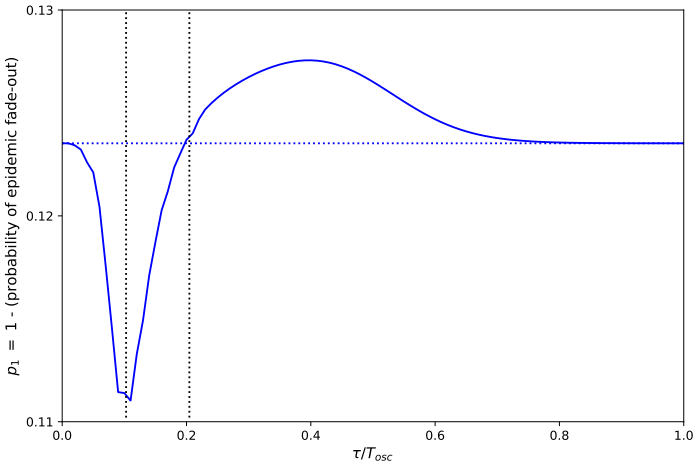


Figure 5: Effect of a short "pulse" of modulation on p_1. SIRS model, $N = 20000$, $\gamma = 52$, $\nu = 1$, $\beta_0 = 156$; and $\beta = 1.1\beta_0$ between times τ and $\tau + T_{osc}/100$, and otherwise $\beta = \beta_0$. The horizontal dotted line shows p_1 when there is a constant $\beta = \beta_0$. The vertical dotted lines correspond respectively to the times of points C and D in Figure 2, if the outbreak follows the deterministic approximation.

the SIRS model with $N = 20000$, $\gamma = 52$, $\nu = 1$, and $\beta_0 = 156$ (corresponding to $R_0 = 3$).

Figure 5 shows two effects. The first effect is that early in the outbreak – more or less between points B and D in Figure 2, when I is large – an increase in β gives a decrease in p_1; and this effect is at its greatest near point C, where I is at its peak. As we have discussed previously [3], this is a slightly counterintuitive effect, but arises due to the fact that a higher β depletes the number of susceptibles, so the outbreak is more likely to "burn out". The magnitude and timing of the effect, including the fact that it peaks near C, is a new result.

The "crossover" point, when increasing β begins to have a positive rather than a negative effect on p_1, is just before point D. This is consistent with our previously reported result [4].

After that point, when the CTMC is in the "trough", the second effect dominates: an increase in β gives an increase in p_1.

When we consider time-dependent modulation models in the following sections, all the results can be explained in terms of a combination of these two effects, to a first order.

4.3. Influenza-like parameters

The influenza virus mutates relatively rapidly, and individuals do not have lifetime immunity. So it is waning immunity, rather than population turnover, that is the primary source of the replenishment of susceptibles. Therefore we model influenza-like parameters with the SIRS model, with a relatively high ν. We consider $\nu = 0.5$ years^{-1} (corresponding to a mean loss of immunity time of 2 years), and $\gamma = 52$ years^{-1} (corresponding to an infectious time of a week) [12].

We model the seasonal variation in β using a cosine wave. We assume β is at its maximum in the middle of winter, and at its minimum in the middle of summer. So for a mean R_0 value of $\overline{R_0}$, and some modulation quantity m,

$$R_0(t) = \overline{R_0}(1 + m\cos(2\pi(t + \phi))),$$

where t is time (in years, from the start of the outbreak) and ϕ is the phase, corresponding to the time from the middle of winter to the start of the outbreak, also in years. Then $\beta(t) = \gamma R_0(t)$.

There is a range of possible values of R_0 for influenza, in some cases $R_0 = 6$ or even higher [23]. Here we are more interested in trends, rather than a specific R_0 or even a specific disease, so we model a range of values of $\overline{R_0}$. Trends were similar across a range of values for N and m. Here we show two representative plots, showing two different values of m: $m = 0.1$ in Figures 6 and $m = 0.2$ in Figure 7, for $N = 20000$. We include, for comparison, one trace for $\nu = 1$ instead of $\nu = 0.5$ in Figure 6.

We also include, as a check, the range of p_1 values from Monte Carlo simulations, for a selection of parameters. These confirm the accuracy of the method.

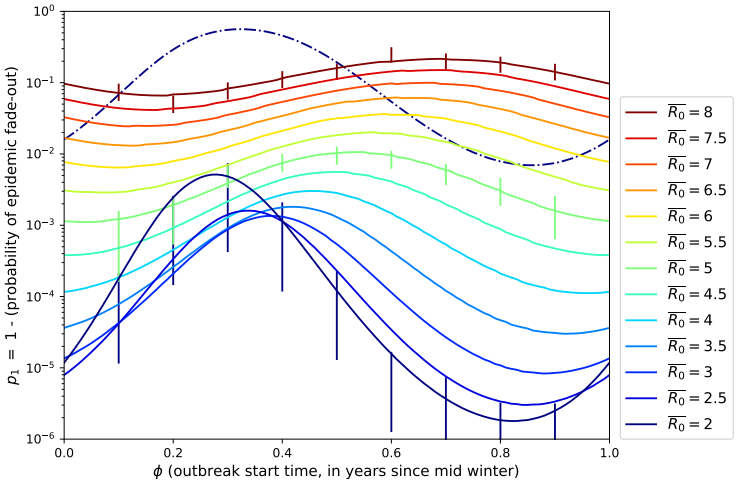


Figure 6: Effect of cosine wave modulation on p_1 for influenza-like parameters: $N = 20000$, $\gamma = 52$, $\nu = 0.5$, and $R_0(t) = \overline{R_0}(1 + 0.1\cos(2\pi(t + \phi)))$. The vertical bars show the range ($\pm 2\sigma$) of values given by Monte Carlo simulations, for $R_0 = 8, 5$ and 2. The "chained" trace shows identical parameters for $R_0 = 2$, except that $\nu = 1$ instead of 0.5.

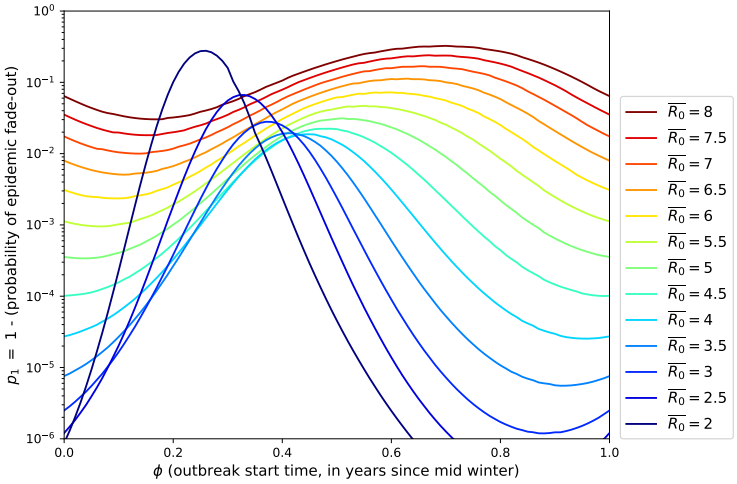


Figure 7: Effect of cosine wave modulation on p_1 for influenza-like parameters: $N = 20000$, $\gamma = 52$, $\nu = 0.5$, and $R_0(t) = \overline{R_0}(1 + 0.2\cos(2\pi(t + \phi)))$.

Setting $\nu = 1$ does not seem to be realistic, but its plot in Figure 6 (the "chained" trace for $R_0 = 2$) serves to illustrate how p_1 has a dramatic dependence on γ/ν – that is, the ratio between the recovery rate and the regeneration rate (or equivalently, the ratio between the mean immunity period and the mean infectious period). A value of $\gamma/\nu = 52$ gives quite high p_1 values, but these fall by two to three orders of magnitude when γ/ν is doubled to 104. So while we may observe trends in p_1 versus $\overline{R_0}$, it is γ/ν – a measure of the regeneration rate – which is far more significant when it comes to determining whether an infection survives after its first wave.

However using $\gamma = 52$ and $\nu = 0.5$, Figures 6 and 7 suggest that the chances of an influenza outbreak surviving after its first wave of infection are low, in the absence of further importation of the virus, unless $\overline{R_0}$ is quite high. This agrees with the observation that there tends to be a single outbreak per season. As previously noted, there have been exceptions such as the 1918-19 Spanish flu [25] and 2009 swine flu [14, 30] pandemics, and these may correspond to $\overline{R_0}$ towards the higher end of these ranges.

A second observation is that the peaks in the graph shift to the right as $\overline{R_0}$ increases. This is due to the fact that T_{osc} is dependent on R_0 (14). This may have implications for determining when is the most critical time of year for an influenza outbreak to occur.

A third observation is that the degree of dependence on the start time varies with $\overline{R_0}$. That is, for low values of $\overline{R_0}$ (such as 2), p_1 varies more dramatically with ϕ; while this seasonal variation is less pronounced for higher values of $\overline{R_0}$.

Furthermore, comparing Figures 6 and 7, we see that this effect becomes more pronounced as the amount of modulation increases. The reason for this is essentially that the minimum value of $R_0(t)$ is important. If $R_0(t)$ is reduced at a critical time in the outbreak cycle (corresponding to the trough at $\tau/T_{osc} \approx 0.1$ in Figure 5) then it causes an increase in p_1. Clearly this effect will be greater, and extend to higher values of $\overline{R_0}$, as the modulation m increases.

In a CTMC with constant β, we previously reported that, if one plots p_1 versus R_0 (by varying β and keeping other parameters constant) there is a local minimum near $R_0 = 2$ [3]. We can now report that, when there is seasonal variation in β, this local minimum is at a somewhat higher mean value of $R_0(t)$, $\overline{R_0}$, and this higher value largely depends on the amount of seasonal variation in $R_0(t)$. In the examples shown, this local minimum p_1 occurs instead at $\overline{R_0} \approx 3$ when the modulation m is 10% (Figure 6), and at $\overline{R_0} \approx 4$ when the modulation is 20% (Figure 7).

4.4. Measles-like parameters

The second test case we modelled was measles-like parameters. These differ from influenza in a number of ways: R_0 is higher, people generally have lifetime immunity, and the population is not naive, with the majority of the population usually immune (having been infected as a child, or been vaccinated).

We used the SIR-with-demography model, with $\mu = 1/70$, corresponding to an average 70-year lifetime [1]. We used an upper value R_0 of 16 [1] and $\gamma = 41$

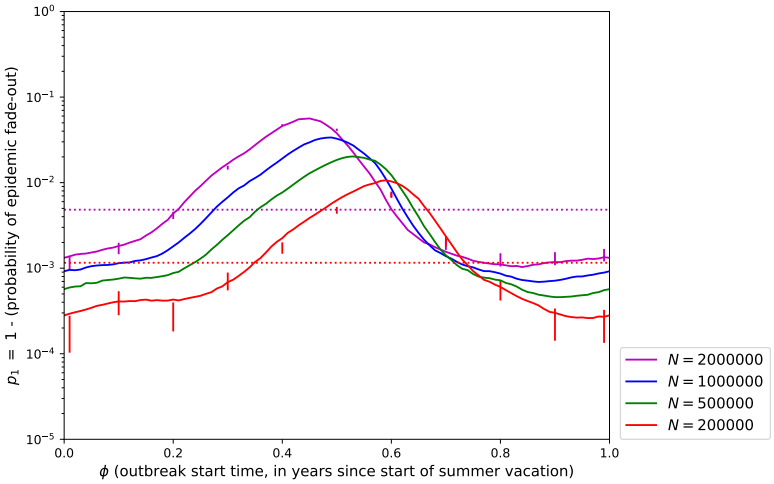


Figure 8: Effect of school vacation modulation on p_1, for measles-like parameters, for different values of N. $\gamma = 41$, $\mu = 1/70$, $\beta_0 = 16(\gamma + \mu)$, and $\beta = 0.8\beta_0$ during the first $6/52$ of the year and three shorter vacations of $2/52$ of the year, otherwise $\beta = \beta_0$, and $(S_0, I_0) = (0.08555N, 1)$. The horizontal dotted lines show p_1 when $\overline{\beta}$, the mean β, is used as a constant ($\overline{\beta} = 15.2615$). The vertical bars show the range ($\pm 2\sigma$) of values given by Monte Carlo simulations, for $N = 200000$ and $N = 2000000$.

(corresponding to an average 9 day infectious period [11]). According to (8), this gives an infection cycle time of approximately 2.1 years, quite close to the 2-year cycle that measles has been observed to follow [13].

We found overall trends did not vary a lot as R_0 varied. This is possibly because all feasible R_0 values for measles are not in the vicinity of $R_0 = 2$. So instead here we present trends as N is varied, since the Critical Community Size (CCS) is a historically interesting number in the study of measles.

To model time dependence in measles, we used school vacations instead of the season to model changes in R_0, with R_0 dropping by 20% during school vacations. To model school vacations, we followed the Australian model of a 6-week summer vacation, then a school year of four 10-week terms with a 2-week break in between each term (apart from summer break). So if ϕ is the start time of the outbreak, in years from the start of summer vacation, then $(\lfloor 52\,(t+\phi)\rfloor \bmod 52)$ is the week of the year; and

$$\beta(t) = \begin{cases} 0.8\beta_0 & \text{if } \lfloor 52\,(t+\phi)\rfloor \bmod 52 \in \{0,1,2,3,4,5,16,17,28,29,40,41\} \\ \beta_0 & \text{otherwise;} \end{cases} \tag{21}$$

where $\beta_0 = R_{0max}(\gamma+\mu)$ and $R_{0max} = 16$.

For the initial condition, although we need to assume that the population is not naive, the exact choice is not important, so long as it is reasonably realistic, because we are looking for trends rather than solving a particular problem. Since the deterministic minimum I value occurs when $S = S_e$, we decided to assume that the previous outbreak died out in the vicinity of $S = \overline{S_e}$, where $\overline{S_e}$ is the S_e value corresponding to $\overline{R_0}$, the mean R_0 value (so $\overline{R_0} = 15.2615$ and $\overline{S_e} = N/\overline{R_0}$). We then assumed 1.5 years of replenishment at rate $\mu(N - \overline{S_e})$. That is, we used as the initial state,

$$S_0 = \overline{S_e} + (1.5)(N - \overline{S_e})\mu = 0.08555N; \quad I_0 = 1.$$

The results for different values of N are shown in Figure 8. As in Figure 6, we include the range of p_1 values from Monte Carlo simulations, for a selection of parameters, as a check that the method is accurate.

The maxima in Figure 8 are due to the timing of the summer vacation. If the summer vacation coincides with the peak of the outbreak, then p_1 is increased. That is, the outbreak infects fewer individuals, so more susceptibles remain after the initial wave, so epidemic fade-out is less likely. According to Figure 5, the optimal time to reduce β is after about $0.1\,T_{osc}$. So with an approximate 2-year cycle, we would expect p_1 to be maximised when the outbreak begins about 0.2 years before this peak; that is, at $\phi \approx 0.8$. In Figure 8, we observe the peaks somewhat earlier than this, because on average the outbreak takes longer to "start up" in the measles model, than in the flu-like model used in Figure 5.

This mean start-up time is also the reason why the timing of the peak varies with population size. An initial infectious number of $I = 1$ is proportionally lower for larger N, so the mean start-up time is longer. So as N increases, the peak p_1 occurs at a lower ϕ value.

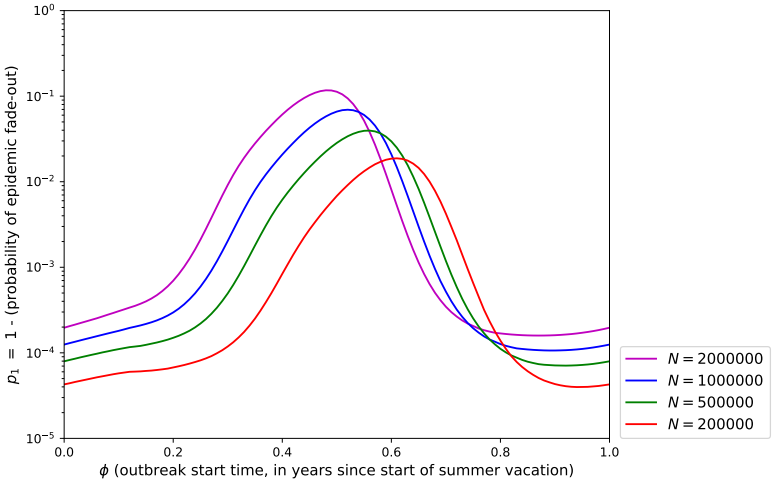


Figure 9: Effect of summer vacation modulation on p_1, for measles-like parameters, for different values of N. This uses identical parameters to Figure 8 ($\gamma = 41$, $\mu = 1/70$, $\beta_0 = 16(\gamma+\mu)$, $(S_0, I_0) = (0.08555N, 1)$) except that $\beta = 0.8\beta_0$ only during the first $6/52$ of the year, otherwise $\beta = \beta_0$.

The shorter vacations have a smaller effect on the general wave shape. Figure 9 shows the effect of using a simpler model where there is only a summer vacation, and the three shorter term breaks are not modelled; that is (21) is modified to:

$$\beta(t) = \begin{cases} 0.8\beta_0 & \text{if } \lfloor 52\,(t+\phi) \rfloor \bmod 52 < 6 \\ \beta_0 & \text{otherwise.} \end{cases} \tag{22}$$

We see that, compared to Figure 8, Figure 9 is "smoother" and has a more pronounced peak, but the overall trends are mostly the same; confirming that the peaks in Figure 8 are due to the timing of the summer vacation.

Let $\overline{\beta}$ be the mean $\beta(t)$ in (21); that is, $\overline{\beta} = (40\beta_0 + 12(0.8\beta_0))/52)$. The dotted lines in Figure 8 show p_1 if β is constant at $\overline{\beta}$, for $N = 200000$ and $N = 2000000$. Though it is not visually obvious because of the logarithmic scale, the mean p_1 over the year is considerably higher using the varying β, than when using the constant $\beta = \overline{\beta}$. So on average, seasonal forcing in β leads to an increase in p_1. That is, seasonal forcing makes epidemic fade-out less likely to occur. So whereas epidemic fade-out might seem almost certain when using time-invariant parameters; the use of time-varying β indicates that there is a significant probability of the infection persisting after the first trough, if the outbreak begins at certain times of the year.

Finally, we note that, even with this increase in p_1 due to seasonal forcing, these values of p_1 are significantly lower than what is generally observed. That is, with the classic CCS for measles in an isolated population being about $200,000$ [6], we would expect $p_1 > 0.5$ for $N \gtrapprox 200,000$, but these results give lower p_1 values than that. We believe this is because the assumption of unstructured homogeneous populations leads to an underestimation of p_1. In reality, pockets of infection (due to structure and non-homogeneity in the population) make the probability of persistence (that is, p_1) higher than our method predicts. So while seasonal variation in a homogeneous population increases p_1, it does not entirely account for the observed values.

5. Conclusion

We have presented a numerical method for calculating p_0, the probability of epidemic fade-out (and its complement p_1), when the transmission rate parameter β is time-dependent. We have done this for two different but related models. We have confirmed that it is fast and accurate.

We have presented results for a number of models of time dependence. For a model which only modifies β for a single block of time (Section 4.2), we see that an increase in β causes a decrease in p_1 during the peak of an infection, while an increase in β has the opposite effect during the trough of an infection. This agrees with previous results.

For influenza-like models (Section 4.3), we observed that p_1 is much more seasonally dependent for low values of R_0 (near 2) than higher values. This seasonal dependence means that outbreaks with low R_0 have a higher probability of persisting than those with slightly higher R_0 values. While with no

modulation there is a local minimum in p_1 in the region of $R_0 = 2$, the point of this local minimum increases as the seasonal variation in R_0 increases, to about $R_0 = 4$ when there is 20% modulation.

For measles-like models (Section 4.4), we see that the timing of the introduction of the infection, relative to the summer vacation, can increase p_1; and that the average p_1 increases, bringing it closer to matching the observed data for measles.

6. Conflicts of interest

None.

7. Acknowledgments

This work is supported by an APA Scholarship (PB), an Australian Research Council Future Fellowship (JVR; FT130100254), and the NHMRC (JVR; CRE PRISM^2).

8. References

[1] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: dynamics and control*. Oxford University Press, Oxford, 1991.

[2] F. G. Ball. The threshold behaviour of epidemic models. *Journal of Applied Probability*, 20(2):227–241, 1983. doi: 10.2307/3213797.

[3] P. G. Ballard, N. G. Bean, and J. V. Ross. The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography. *Journal of Theoretical Biology*, 393:170–178, 2016. doi: 10.1016/j.jtbi.2016.01.012.

[4] P. G. Ballard, N. G. Bean, and J. V. Ross. Epidemic intervention to maximise the probability of epidemic fade-out. *Mathematical Biosciences*, 293:1–10, November 2017. doi: 10.1016/j.mbs.2017.08.003.

[5] M. S. Bartlett. Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 81–109. University of California Press, Berkeley, 1956.

[6] M. S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Society: Series A*, 120(1):48–70, 1957. doi: 10.2307/2342553.

[7] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith. Cython: The best of both worlds. *Computing in Science & Engineering*, 13(2):31–39, 2011. doi: 10.1109/MCSE.2010.118.

[8] T. Britton. Stochastic epidemic modelling and analysis: current perspective and future challenges (Part 2), at *Infectious Disease Dynamics*, Isaac Newton Institute for Mathematical Sciences, August 2013. http://www.newton.ac.uk/seminar/20130820163017001.

[9] T. Britton, T. House, A. L. Lloyd, D. Mollison, S. Riley, and P. Trapman. Five challenges for stochastic epidemic models involving global transmission. *Epidemics*, 10:54–57, 2015. doi: 10.1016/j.epidem.2014.05.002.

[10] A. J. K. Conlan and B. T. Grenfell. Seasonality and the persistence and invasion of measles. *Proceedings of the Royal Society B*, 274:1133–1141, 2007. doi: 10.1098/rspb.2006.0030.

[11] Department of Health and Human Services, Victoria, Australia. Measles, 2015. https://www2.health.vic.gov.au/public-health/infectious-diseases/disease-information-advice/measles, Accessed: 2017-09-05.

[12] Department of Health, Australian Government. Influenza Infection, CDNA National Guidelines For Public Health Units, August 2011. http://www.health.gov.au/internet/main/publishing.nsf/Content/cdna-song-influenza.htm, Accessed: 2018-01-04.

[13] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Sons, Chichester, 2000.

[14] L. J. Donaldson, P. D. Rutter, B. M. Ellis, F. E. C. Greaves, O. T. Mytton, R. G. Pebody, and I. E. Yardley. Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study. *BMJ*, 339:b5213, 2009. doi: 10.1136/bmj.b5213.

[15] S. N. Ethier and T. G. Kurtz. *Markov Processes*. John Wiley & Sons, Hoboken, 1986.

[16] P. E. M. Fine and J. A. Clarkson. Measles in England and Wales – I: An Analysis of Factors Underlying Seasonal Patterns. *International Journal of Epidemiology*, 11(1):5–14, 1982. doi: 10.1093/ije/11.1.5.

[17] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976. doi: 10.1016/0021-9991(76)90041-3.

[18] B. Grenfell and J. Harwood. (Meta)population dynamics of infectious diseases. *Trends in Ecology and Evolution*, 12(10):395–399, October 1997. doi: 10.1016/S0169-5347(97)01174-9.

[19] A. Kamenev and B. Meerson. Extinction of an infectious disease: A large fluctuation in a nonequilibrium system. *Physical Review E*, 77:061107, 2008. doi: 10.1103/PhysRevE.77.061107.

[20] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, New Jersey, 2008.

[21] E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley & Sons, New York, 8th edition, 1999.

[22] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970. doi: 10.2307/3212147.

[23] J. D. Mathews, C. T. McCaw, J. McVernon, E. S. McBryde, and J. M. McCaw. A biological model for influenza transmission: Pandemic planning implications of asymptomatic infection and immunity. *PLoS ONE*, 2(11): e1220, 2007. doi: 10.1371/journal.pone.0001220.

[24] B. Meerson and P. V. Sasorov. WKB theory of epidemic fade-out in stochastic populations. *Physical Review E*, 80:041130, 2009. doi: 10.1103/PhysRevE.80.041130.

[25] D. M. Morens, J. K. Taubenberger, H. A. Harvey, and M. J. Memoli. The 1918 influenza pandemic: Lessons for 2009 and the future. *Critical Care Medicine*, 38(4 (Suppl)):e10–e20, 2010. doi: 10.1097/CCM.0b013e3181ceb25b.

[26] I. Nåsell. On the time to extinction in recurrent epidemics. *Journal of the Royal Statistical Society: Series B*, 61:309–330, 1999. URL `http://www.jstor.org/stable/2680643`.

[27] I. Nåsell. Stochastic models of some endemic infections. *Mathematical Biosciences*, 179:1–19, 2002. doi: 10.1016/S0025-5564(02)00098-6.

[28] P. K. Pollett. On a model for interference between searching insect parasites. *The Journal of the Australian Mathematical Society, Series B*, 32: 133–150, 1990. doi: 10.1017/S0334270000008390.

[29] P. K. Pollett. Diffusion approximations for ecological models. In *Proceedings of the International Congress on Modelling and Simulation*, volume 2, pages 843–848. Modelling and Simulation Society of Australia and New Zealand, 2001.

[30] R. Reintjes, E. Das, C. Klemm, J. H. Richardus, Verena Keßler, and A. Ahmad. "Pandemic Public Health Paradox": Time Series Analysis of the 2009/10 Influenza A / H1N1 Epidemiology, Media Attention, Risk Perception and Public Reactions in 5 European Countries. *PLoS ONE*, 11(3): e0151258, 2015. doi: 10.1371/journal.pone.0151258.

[31] Z. Patterson Ross, N. Komadina, Y-M Deng, N. Spirason, H. A. Kelly, S. G. Sullivan, I. G. Barr, and E. C. Holmes. Inter-seasonal influenza is characterized by extended virus transmission and persistence. *PLoS Pathogens*, 11(6):e1004991, 2015. doi: 10.1371/journal.ppat.1004991.

[32] O. A. van Herwaarden. Stochastic epidemics: the probability of extinction of an infectious disease at the end of a major outbreak. *Journal of Mathematical Biology*, 35:793–813, 1997. doi: 10.1007/s002850050077.

[33] R. Yaari, G. Katriel, A. Huppert, J. B. Axelsen, and L. Stone. Modelling seasonal influenza: the role of weather and punctuated antigenic drift. *Journal of the Royal Society Interface*, 10:20130298, 2013. doi: 10.1098/rsif.2013.0298.

6 Software

An important part of the work was the choice of software to use for simulations and calculations. The three main choices were Matlab, Julia and Cython.

Matlab [45] is designed for scientific use and, of the three, probably has the most complete set of mathematical libraries. However it has (in our opinion) a relatively clumsy syntax compared to the other two options. Also, unlike Julia and Cython, it is commercial software, and that limited how much it could be used off-site.

Julia is a relatively new language, first released in 2012 [16]. Its syntax is quite similar to Matlab, but avoids many of Matlab's idiosyncrasies, and has the advantage that it is free. One main drawback was that, at least when this work began in 2014, its collection of libraries was quite small.

Cython is a third party enhancement to the Python language [14], and is also free. Python is a very popular general purpose language, and there exist well supported third party packages for scientific and mathematical functions, called NumPy [58] and SciPy [33]. Python itself is not a very fast language, but Cython considerably speeds Python code up, if the user adds a few well chosen directives to the critical parts of the program.

Numba, which is another package to accelerate the speed of Python [40], was also considered, and on one test case it ran faster than any of Matlab, Julia and Cython. But at the time of testing, it was incompatible with some parts of SciPy. This made it unusable, so it had to be rejected.

Between the other three, the choice for each stage of the project was based on performance, as detailed below.

6.1 Paper 1

An early version of the numerical algorithms used in Paper 1 [6] was run in Matlab, Julia and Cython, with the results shown in Tables 3 and 4.

Population size (N)	Cython	Julia	Matlab
1,000	3	65	15
2,000	9	not tested	60

Table 3: Evaluation times in seconds, for an early version of the "exact model" in Paper 1 [6, Section 3.3.1].

Population size (N)	Cython	Julia	Matlab
200,000	5	75	13
500,000	13	526	33
1,000,000 (case a)	38	2200	105
1,000,000 (case b)	356	15850	805

Table 4: Evaluation times in seconds, for an early version of the "approximate model" in Paper 1 [6, Section 3.3.2].

As a result of these tests, it was concluded that Cython was both the fastest, and the one which was able to handle the largest data arrays (and hence, could simulate up to the largest population size). Therefore all algorithms were written in Cython.

For random (Monte Carlo) simulations (which were run as a check on some of the results), Cython and Julia were of comparable speed, and somewhat

faster than Matlab. Cython was chosen over Julia as a matter of convenience, so that all tests could be run in the same environment.

6.2 Paper 2

As noted at the start of Section 3.3 of Paper 2 [7], the Markov decision process calculations described in this paper required very large probability transition matrices: a population size N has a truncated state space of approximately N^2, requiring an approximately $N^2 \times N^2$ matrix. These matrices are also very sparse: each state has transitions to at most 4 other states, so each row has at most five non-zero entries.

Matlab, Julia and Cython (via the SciPy package) all have provision to efficiently represent such matrices, with data structures known as sparse arrays. However the SciPy implementation was quite complicated and so, in the interests of keeping the code simple, Cython was eliminated from consideration.

So Matlab and Julia were tested for speed, with the results in Table 5.

Population size (N)	Julia	Matlab
100	1	3
300	15	57
600	383	4990
800	983	test abandoned

Table 5: Evaluation times in seconds, for solving the Linear Program algorithm in Section 4.2 of Paper 2 [7]. For a population size of N, the truncated state space is about N^2, so the Markov decision process algorithms require a sparse array of about size $N^2 \times N^2$.

Based on these results, it appeared that Julia had the more efficient implementation of sparse arrays. Therefore Julia was used for the Markov decision process evaluations in Sections 3.2, 4.2 and 5.1 of Paper 2.

The p_0 evaluations (in Sections 5.2 and 5.3) used the approximate algorithm of Paper 1 [6, Section 3.3.2], which was written in Cython.

6.3 Paper 3

The slowest part of the algorithm in Paper 3 was the semi-continuous approximation described in [8, Section 3.2]. This was designed to work with any DE (differential equation) solver, so this was originally implemented in Matlab, whose "ode45" DE solver [46] was more robust than anything available in Python or Julia.

However, for very large population sizes (over $N = 100,000$), this began to run out of memory and run extremely slowly. On the other hand, a simple Euler method DE solver ran comfortably up to the $N = 2,000,000$ used in the paper. This was implemented in Cython, because the code was related to the code used in [6, Section 3.3.2], which had been shown in Table 4 to perform best in Cython. So Cython was used for the results presented in Paper 3 [8].

As with the Paper 1 [6], all random simulations were also run in Cython.

6.4 Summary

So in summary, almost all algorithms were written in Cython, because that was tested to run the fastest and handle the largest population sizes. The only exceptions were the Markov decision processes in Paper 2, which were written in Julia, because Julia had the best handling of large sparse arrays.

All plotting was done in Python, using the Python Matplotlib package [32].

6.5 Software online

A tar file has been created containing the source code for the algorithms described in these papers. It is available online at:

`https://figshare.com/s/b2c0f0adf76ff078448b`

7 Conclusion

We have presented three papers on the subject of the probability of epidemic fade-out, particularly in the Markovian SIR-with-demography model, and we believe we have made significant contributions in a number of areas.

First, we provided a definition of p_0, the probability of epidemic fade-out [6, Section 3.1], which appears not to have been done before.

In Paper 1 [6], we devised a novel, efficient, numerical evaluation method for p_0. We verified that it was more accurate than previously published estimates, and faster than other methods.

This allowed p_0 to be tested across a range of parameters. As a result of that testing, we made the unexpected observation that p_0 has a non-monotonic relationship with β, the transmission rate parameter; and that p_0 has a local maximum in the region of $R_0 \approx 2$. We also provided an explanation for this. This explanation tied with previous observations that epidemic fade-out is related to the depletion of susceptibles, and we explained why this effect is maximised in the region of $R_0 \approx 2$.

In Paper 2 [7], we used this non-monotonicity, and showed how to maximise p_0 if an intervention to control β is available. We showed how an almost optimal policy can be distilled to a single formula – reduce β when $(\gamma+\mu)I < \mu(N-S)$ – which is both simple, and independent of the values of β. We also showed that this policy is robust, in that a close-to-maximum p_0 is obtained even if this policy is only followed approximately.

Then in Paper 3 [8], we extended the techniques of the first paper, to account for time dependence in β; and applied this to influenza-like and measles-like parameters. In particular, we noted that the local maximum in p_0 occurs at a larger value of R_0, perhaps in the region of $R_0 \approx 4$ for realistic amounts

of seasonal variation.

For future work, it would be good to extend both the method of calculating p_0, and the properties of p_0, to more complex models. For instance, both the inclusion of an "exposed" state (the SEIR model), and the modelling of the infectious time with an Erlang distribution, produce more realistic epidemic models [61]. These models have state spaces with more than two dimensions, so it is likely that further approximations would be required, to extend the techniques used in Papers 1 [6] and 3 [8]. If such extensions cannot be found, it should still be possible to examine these models using Monte Carlo simulations.

It should also be possible to extend the optimisation work of Paper 2 [7]. There are at least two directions this may take. One would be to extend it to more complex models such as those described above. It should be possible to modify the almost-optimal policy formula for these models.

Another possibility would be to optimise a quantity other than p_0, such as the final epidemic size, with or without with some form of time discounting. That is, are there circumstances under which one can withhold an intervention (such as antivirals) and have the effect of reducing the final epidemic size? This could have implications for the optimal response to an infectious outbreak. Or even in cases where treatment cannot be ethically withheld, the techniques could be used to analyse the reasons behind the size of an outbreak.

In conclusion, we have used mathematical, computational and modelling techniques to extend our knowledge about epidemic fade-out. We believe these analyses have enhanced the understanding of the processes behind epidemic fade-out, and how to control it. We also believe it lays the ground for future work in this area, towards the ultimate goal of people knowing how best to respond to outbreaks of infectious diseases.

8 References

[1] R. M. Anderson. Discussion: The Kermack-McKendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53:3–32, 1991. doi: 10.1007/BF02464422.

[2] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: dynamics and control*. Oxford University Press, Oxford, 1991.

[3] J. R. Artalejo and M. J. Lopez-Herrero. Quasi-stationary and ratio of expectations distributions: A comparative study. *Journal of Theoretical Biology*, 266:264–274, 2010. doi: 10.1016/j.jtbi.2010.06.030.

[4] N. T. J. Bailey. A simple stochastic epidemic. *Biometrika*, 37(3/4): 193–202, 1950. URL `http://www.jstor.org/stable/2332371`.

[5] F. G. Ball. The threshold behaviour of epidemic models. *Journal of Applied Probability*, 20(2):227–241, 1983. doi: 10.2307/3213797.

[6] P. G. Ballard, N. G. Bean, and J. V. Ross. The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography. *Journal of Theoretical Biology*, 393:170–178, 2016. doi: 10.1016/j.jtbi.2016.01.012.

[7] P. G. Ballard, N. G. Bean, and J. V. Ross. Intervention to maximise the probability of epidemic fade-out. *Mathematical Biosciences*, 293:1–10, November 2017. doi: 10.1016/j.mbs.2017.08.003.

[8] P. G. Ballard, N. G. Bean, and J. V. Ross. The impact of time dependent transmission rate on the probability of epidemic fade-out. Submitted to Journal of Theoretical Biology, 2018.

[9] S. Ballesteros, E. Vergu, and B. Cazelles. Influenza A gradual and epochal evolution: Insights from simple models. *PLoS ONE*, 4(10): e7426, October 2009. doi: 10.1371/journal.pone.0007426.

[10] M. S. Bartlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society: Series B*, 11(2):211–229, 1949. URL `http://www.jstor.org/stable/2984077`.

[11] M. S. Bartlett. Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 81–109. University of California Press, Berkeley, 1956.

[12] M. S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Society: Series A*, 120(1):48–70, 1957. doi: 10.2307/2342553.

[13] M. S. Bartlett. The critical community size for measles in the United States. *Journal of the Royal Statistical Society: Series A*, 123(1):37–44, 1960. URL `http://www.jstor.org/stable/2343186`.

[14] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith. Cython: The best of both worlds. *Computing in Science & Engineering*, 13(2):31–39, 2011. doi: 10.1109/MCSE.2010.118.

[15] C. Bender and S. Orszag. *Advanced Mathematical Method for Scientists and Engineers*. McGraw-Hill, 1978.

[16] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman. Julia: A fast dynamic language for technical computing. arXiv:1209.5145 [cs.PL], 2012. URL `https://arxiv.org/abs/1209.5145`.

[17] L. Brillouin. Remarques sur la méchanique ondulatoire. *Journal de Physique et le Radium*, 7:353–368, 1926.

[18] T. Britton. Stochastic epidemic modelling and analysis: current perspective and future challenges (Part 2), at *Infectious Disease Dynamics*, Isaac Newton Institute for Mathematical Sciences, August 2013. http://www.newton.ac.uk/seminar/20130820163017001.

[19] T. Britton, T. House, A. L. Lloyd, D. Mollison, S. Riley, and P. Trapman. Five challenges for stochastic epidemic models involving global transmission. *Epidemics*, 10:54–57, 2015. doi: 10.1016/j.epidem.2014.05.002.

[20] A. Camacho, S. Ballesteros, A. L. Graham, F. Carrat, O. Ratmann, and B. Cazelles. Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proceedings of the Royal Society B*, 278:3635–3643, 2011. doi: 10.1098/rspb.2011.0030.

[21] A. J. K. Conlan, P. Rohani, A. L. Lloyd, M. Keeling, and B. T. Grenfell. Resolving the impact of waiting time distributions on the persistence of measles. *Journal of the Royal Society Interface*, 7:623–640, 2010. doi: 10.1098/rsif.2009.0284.

[22] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Sons, Chichester, 2000.

[23] P. E. M. Fine and J. A. Clarkson. Measles in England and Wales – I: An Analysis of Factors Underlying Seasonal Patterns. *International Journal of Epidemiology*, 11(1):5–14, 1982. doi: 10.1093/ije/11.1.5.

[24] L. Q. Gao and H. W. Hethcote. Disease transmission models with density-dependent demographics. *Journal of Mathematical Biology*, 30: 717–731, 1992. doi: 10.1007/bf00173265.

[25] C. W. Gardiner. *Handbook of Stochastic Methods*. Springer-Verlag, Berlin, 2nd edition, 1985.

[26] A. E. Gill. *Atmosphere-ocean dynamics*. Academic Press, New York, 1982.

[27] G. Green. On the motion of waves in a variable canal of small depth and width. *Transactions of the Cambridge Philosophical Society*, 6:457–62, 1837.

[28] M. Greenwood. On the statistical measure of infectiousness. *The Journal of Hygiene*, 31(3):336–351, July 1931. URL `http://www.jstor.org/stable/3859298`.

[29] B. Grenfell and J. Harwood. (Meta)population dynamics of infectious diseases. *Trends in Ecology and Evolution*, 12(10):395–399, October 1997. doi: 10.1016/S0169-5347(97)01174-9.

[30] H. W. Hethcote. Qualitative analyses of communicable disease models. *Mathematical Biosciences*, 28(3):335–356, 1976. doi: 10.1016/0025-5564(76)90132-2.

[31] H. W. Hethcote. The mathematics of infectious disease. *SIAM Review*, 42(4):599–653, 2000. URL `http://www.jstor.org/stable/2653135`.

[32] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[33] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. http://www.scipy.org/.

[34] A. Kamenev and B. Meerson. Extinction of an infectious disease: A large fluctuation in a nonequilibrium system. *Physical Review E*, 77: 061107, 2008. doi: 10.1103/PhysRevE.77.061107.

[35] S. Karlin. *A First Course in Stochastic Processes*. Academic Press, New York, 1966.

[36] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927. URL `http://www.jstor.org/stable/94815`.

[37] D. A. Kessler and N. M. Shnerb. Extinction rates for fluctuation-induced metastabilities: A real-space WKB approach. *Journal of Statistical Physics*, 127(5):861–886, 2007. doi: 10.1007/s10955-007-9312-2.

[38] H. A. Kramers. Wellenmechanik und halbzahlige quantisierung. *Zeitschrift für Physik*, 39:828–840, 1926.

[39] R. Kuske, L. F. Gordillo, and P. Greenwood. Sustained oscillations via coherence resonance in SIR. *Journal of Theoretical Biology*, 245:459–469, 2007. doi: 10.1016/j.jtbi.2006.10.029.

[40] S. L. Lam, A. Pitrou, and S. Seibert. Numba: a LLVM-based Python JIT compiler. In *LLVM '15 Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, Austin, November 2015. ACM. doi: 10.1145/2833157.2833162.

[41] E. G. Leigh. The average lifetime of a population in a varying environment. *Journal of Theoretical Biology*, 90(2):213–239, 1981. doi: 10.1016/0022-5193(81)90044-8.

[42] J. Liouville. Sur le développement des fonctions ou parties de fonotions en séries... *Journal de Mathématiques Pures et Appliquées*, 2:16–35, 1827.

[43] J. O. Lloyd-Smith, P. C. Cross, C. J. Briggs, M. Daugherty, W. M. Getz, J. Latto, M. S. Sanchez, A. B. Smith, and A. Swei. Should we expect population thresholds for wildlife disease? *Trends in Ecology and Evolution*, 20(9):511–519, September 2005. doi: 10.1016/j.tree.2005.07.004.

[44] A. J. Lotka. Martini's equations for the epidemiology of immunising diseases. *Nature*, 111(2793):633–634, May 1923.

[45] Mathworks Inc. Matlab, 2018. https://mathworks.com/products/matlab.html.

[46] Mathworks Inc. Matlab ode45, 2018. https://au.mathworks.com/help/matlab/ref/ode45.html.

[47] A. G. McKendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 14:98–130, 1926. doi: 10.1017/S0013091500034428.

[48] B. Meerson and P. V. Sasorov. WKB theory of epidemic fade-out in stochastic populations. *Physical Review E*, 80:041130, 2009. doi: 10.1103/PhysRevE.80.041130.

[49] I. Nåsell. The quasi-stationary distribution of the closed endemic SIS model. *Advances in Applied Probability*, 28(3):895–932, September 1996. doi: 10.2307/1428186.

[50] I. Nåsell. On the time to extinction in recurrent epidemics. *Journal of the Royal Statistical Society: Series B*, 61:309–330, 1999. URL `http://www.jstor.org/stable/2680643`.

[51] I. Nåsell. Stochastic models of some endemic infections. *Mathematical Biosciences*, 179:1–19, 2002. doi: 10.1016/S0025-5564(02)00098-6.

[52] I. Nåsell. A new look at the critical community size for childhood infections. *Theoretical Population Biology*, 67(3):203–216, 2005. doi: 10.1016/j.tpb.2005.01.002.

[53] F. W. J. Olver. Error bounds for the Liouville-Green (or WKB) approximation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57:790–810, 1961. doi: 10.1017/S0305004100035945.

[54] O. Ovaskainen. The quasistationary distribution of the stochastic logistic model. *Journal of Applied Probability*, 38(4):898–907, December 2001. URL `http://www.jstor.org/stable/3215772`.

[55] O. Ovaskainen and B. Meerson. Stochastic models of population extinction. *Trends in Ecology and Evolution*, 25(11):643–652, November 2010. doi: 10.1016/j.tree.2010.07.009.

[56] H. E. Soper. The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society*, 92(1):34–73, 1929. URL `http://www.jstor.org/stable/2341437`.

[57] J. Swinton, M. E. J. Woolhouse, M. E. Begon, A. P. Dobson, E. Ferroglio, B. T. Grenfell, V. Gubertid, R. S. Hails, J. A. P. Heesterbeek, A. Lavazza, M. G. Roberts, P. J. White, and K. Wilson. Microparasite transmission and persistence. In P. J. Hudson, editor, *The Ecology of Wildlife Diseases*, chapter 5, pages 83–101. Oxford University Press, 2002.

[58] S. van der Walt, S. C. Colbert, and G. Varoquaux. The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, March-April 2011. doi: 10.1109/MCSE.2011.37.

[59] O. A. van Herwaarden. Stochastic epidemics: the probability of extinction of an infectious disease at the end of a major outbreak. *Journal of Mathematical Biology*, 35:793–813, 1997. doi: 10.1007/s002850050077.

[60] O. A. van Herwaarden and J. Grasman. Stochastic epidemics: major outbreaks and the duration of the endemic period. *Journal of Mathematical Biology*, 33:581–601, 1995. doi: 10.1007/bf00298644.

[61] H. J. Wearing, P. Rohani, and M. J. Keeling. Appropriate models for the management of infectious diseases. *PLoS Medicine*, 2(7):e174, 2005. doi: 10.1371/journal.pmed.0020174.

[62] G. Wentzel. Eine verallgemeinerung der quantenbedingungen für die zwecke der wellenmechanik. *Zeitschrift für Physik*, 38:518–529, 1926.

[63] Y. Xiao, D. Clancy, N. P. French, and R. G. Bower. A semi-stochastic model for salmonella infection in a multi-group herd. *Mathematical Biosciences*, 200:214–233, 2006. doi: 10.1016/j.mbs.2006.01.006.