INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Hempel K, Earn DJD. 2020 Estimating epidemic coupling between populations from the time to invasion. *J. R. Soc. Interface* **17**: 20200523. http://dx.doi.org/10.1098/rsif.2020.0523

Received: 2 July 2020
Accepted: 3 November 2020

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biomathematics

Keywords:
epidemiological model, contact matrix, spatial mixing, time to invasion, parameter estimation, infectious disease

Author for correspondence:
Karsten Hempel
e-mail: karsten.hempel@gmail.com

THE ROYAL SOCIETY PUBLISHING

Estimating epidemic coupling between populations from the time to invasion

Karsten Hempel and David J. D. Earn

Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1

KH, 0000-0003-3273-4247; DJDE, 0000-0002-7562-1341

Identifying the mechanisms by which diseases spread among populations is important for understanding and forecasting patterns of epidemics and pandemics. Estimating transmission coupling among populations is challenging because transmission events are difficult to observe in practice, and connectivity among populations is often obscured by local disease dynamics. We consider the common situation in which an epidemic is seeded in one population and later spreads to a second population. We present a method for estimating transmission coupling between the two populations, assuming they can be modelled as *susceptible–infected–removed* (SIR) systems. We show that the strength of coupling between the two populations can be estimated from the time taken for the disease to invade the second population. Confidence in the estimate is low if only a single invasion event has been observed, but is substantially improved if numerous independent invasion events are observed. Our analysis of this simplest, idealized scenario represents a first step toward developing and verifying methods for estimating epidemic coupling among populations in an ever-more-connected global human population.

1. Introduction

Mechanistic mathematical models are powerful tools for understanding and predicting how infectious diseases spread in human populations [1–5]. The spread of infections in well-mixed populations has been extensively studied, and continuing research is tackling the effects of seasonal forcing [6–8], intensity and duration of infectiousness [9–14], and contact network structure [15–18].

One area of research that is important for public health policy is forecasting the spatial spread of diseases, which can be greatly advanced by improving estimates of model parameters from real-world data. Estimating parameters of spatial epidemic models is especially difficult [19–21], even for the well studied, highly idealized class of meta-population models [18,22–31]. Here, we consider the simplest meta-population consisting of individuals who reside in one of two 'habitat patches' (e.g. cities). We suppose an epidemic begins in one patch, and we attempt to estimate the degree of spatial coupling to the population in the second patch. In this situation, we investigate whether we can successfully estimate the magnitude of coupling using the observed time taken for the second patch to be infected (the *time to invasion*, t_{inv}).

The specific meta-population model that we use is a two-patch SIR model (§2). We consider both deterministic and stochastic versions of this model (§2) and show that the distribution of times to invasion can be approximated analytically from model parameters (§3.1). We then show how, in the presence of stochasticity, the degree of coupling can be estimated using a maximum-likelihood approach based on one or more observations of t_{inv} (§3.4).

© 2020 The Author(s) Published by the Royal Society. All rights reserved.

2. Two-population susceptible–infected–removed model

In the absence of coupling, we assume that disease dynamics in each patch evolve according to the standard SIR model,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta S\frac{I}{N}, \tag{2.1a}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta S\frac{I}{N} - \gamma I, \tag{2.1b}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I. \tag{2.1c}$$

The three state variables represent the numbers of individuals who are susceptible to infection (S), currently infected and infectious (I), and recovered and removed (R). The total population size, $N = S + I + R$, is necessarily constant (since $\mathrm{d}N/\mathrm{d}t = 0$). The two disease parameters are the rate of transmission (β) and the rate at which infected individuals recover or die (γ). The ***force of infection*** is

$$\Lambda = \beta\frac{I}{N}. \tag{2.2}$$

The ***basic reproduction number***, the average number of secondary cases that result from a single primary case in a completely susceptible population [1], is

$$\mathcal{R}_0 = \frac{\beta}{\gamma}. \tag{2.3}$$

If we take the time unit to be the mean infectious period ($1/\gamma$) then $\mathcal{R}_0$ is the only disease parameter. Implicit in equation (2.1) are assumptions that recovered individuals remain immune permanently and that vital dynamics (births and deaths) can be ignored (both these assumptions are reasonable for most infectious diseases on the timescale of invasion that concerns us here). In addition, the population in any given patch is assumed to be homogeneously mixed.

2.1. Form of transmission coupling

We assume that coupling of disease dynamics between the two patches arises because residents of one patch sometimes visit the other patch temporarily. We model this with a ***coupling matrix*** $c = (c_{ij})$, where c_{ij} is the proportion of the residents of patch j visiting patch i at any time.[1] Since we are considering only two patches, and the entries are proportions, the most general coupling matrix is

$$c = \begin{pmatrix} 1-m_1 & m_2 \\ m_1 & 1-m_2 \end{pmatrix}, \tag{2.4}$$

where $0 \le m_i \le 1$. The ***coupling parameters*** m_i are proportions, which can be interpreted in various ways (e.g. the proportion of time that individuals typically spend outside their home region, or the proportion of individuals that are visiting the other patch at any time). Note that with only two patches, if the focal patch is i then the other patch is $j = 3 - i$. Thus, using subscripts on state variables to identify *populations* (i.e. the patches in which individuals are *resident*), the number of individuals in patch i at any time is

$$(1-m_i)N_i + m_jN_j, \quad i = 1, 2,\ j = 3-i, \tag{2.5}$$

and the number of those that are currently infected is

$$(1-m_i)I_i + m_jI_j, \quad i = 1, 2,\ j = 3-i. \tag{2.6}$$

The force of infection on *residents* of patch i arises from interactions that occur in both patches. For the $(1-m_i)S_i$ susceptibles who are resident in patch i and currently located in patch i, the force of infection is

$$\beta\frac{(1-m_i)I_i + m_jI_j}{(1-m_i)N_i + m_jN_j}, \quad i = 1, 2,\ j = 3-i. \tag{2.7}$$

whereas the force of infection on the m_iS_i susceptible residents of patch i who are currently in patch j is

$$\beta\frac{m_iI_i + (1-m_j)I_j}{m_iN_i + (1-m_j)N_j}, \quad i = 1, 2,\ j = 3-i. \tag{2.8}$$

The total force of infection on residents of patch i is the sum of these two contributions, namely

$$\Lambda_i = \beta\left[(1-m_i)\frac{(1-m_i)I_i + m_jI_j}{(1-m_i)N_i + m_jN_j} + m_i\frac{m_iI_i + (1-m_j)I_j}{m_iN_i + (1-m_j)N_j}\right] \quad i = 1, 2,\ j = 3-i. \tag{2.9}$$

This formulation avoids the need to explicitly model the movements of individuals among populations (as is sometimes done [26]).

2.2. Deterministic model

Our two-population model is, for $i = 1, 2$,

$$\frac{\mathrm{d}S_i}{\mathrm{d}t} = -S_i\Lambda_i, \tag{2.10a}$$

$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = S_i\Lambda_i - \gamma I_i, \tag{2.10b}$$

$$\frac{\mathrm{d}R_i}{\mathrm{d}t} = \gamma I_i, \tag{2.10c}$$

where Λ_i is defined in equation (2.9) and the (constant) size of each population is $N_i = S_i + I_i + R_i$ for $i = 1, 2$.

If all individuals are initially susceptible and a resident of patch i is infected then an epidemic will occur (in population i) if the number of cases in population i is initially increasing, i.e. if $\mathrm{d}I_i/\mathrm{d}t > 0$ in the limit that $S_i \to N_i$ and $I_i \to 0$ (given $S_j = N_j$ and $I_j = 0$). Retaining the notation $\mathcal{R}_0$, as in equation (2.3), for the basic reproduction number of the uncoupled model ($m_1 = m_2 = 0$), and defining $\mathcal{R}_{i,j}$ via

$$\mathcal{R}_{i,i} = \mathcal{R}_0\left[\frac{(1-m_i)^2N_i}{(1-m_i)N_i + m_jN_j} + \frac{m_i^2N_i}{m_iN_i + (1-m_j)N_j}\right] \tag{2.11a}$$

and

$$\mathcal{R}_{i,j} = \mathcal{R}_0\left[\frac{(1-m_i)m_jN_i}{(1-m_i)N_i + m_jN_j} + \frac{m_i(1-m_j)N_i}{m_iN_i + (1-m_j)N_j}\right], \tag{2.11b}$$

we can rewrite equation (2.10*b*)

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \left(\begin{bmatrix} \mathcal{R}_{1,1} & \mathcal{R}_{1,2} \\ \mathcal{R}_{2,1} & \mathcal{R}_{2,2} \end{bmatrix}\gamma - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\gamma\right)\begin{pmatrix} I_1 \\ I_2 \end{pmatrix}, \tag{2.12}$$

from which it follows that the next-generation matrix [37,38] is

$$\begin{bmatrix} \mathcal{R}_{1,1} & \mathcal{R}_{1,2} \\ \mathcal{R}_{2,1} & \mathcal{R}_{2,2} \end{bmatrix}. \tag{2.13}$$

The spectral radius of this matrix, i.e. the basic reproduction number of the two-patch system, is

$$\rho = \frac{\mathcal{R}_{1,1} + \mathcal{R}_{2,2}}{2} + \sqrt{\mathcal{R}_{1,2}\mathcal{R}_{2,1} + (\mathcal{R}_{1,1} - \mathcal{R}_{2,2})^2}. \tag{2.14}$$

Our further analysis in this paper applies to the general case considered above (i.e. it is possible that $N_1 \neq N_2$ and $m_1 \neq m_2$). However, we note that in the special case that $N_1 = N_2$ and $m_1 = m_2$ ($\equiv m$), equation (2.11) reduces to

$$\mathcal{R}_{i,i} = \mathcal{R}_0[1 - 2m(1 - m)] \tag{2.15a}$$

and

$$\mathcal{R}_{i,j} = \mathcal{R}_0 2m(1 - m), \tag{2.15b}$$

and the spectral radius (2.14) simplifies to

$$\rho = \mathcal{R}_0, \tag{2.16}$$

i.e. the basic reproduction number of the two-patch system is the same as that of the single patch system. In this case, there is a simple partitioning of $\mathcal{R}_0$:

$$\mathcal{R}_0 = \mathcal{R}_{i,i} + \mathcal{R}_{i,j}. \tag{2.17}$$

In fact, equation (2.17) holds true if $m_1 \neq m_2$, requiring only that $N_1 = N_2$. In addition, note that

$$\mathcal{R}_{i,j} = \mathcal{R}_{i,i} - (1 - 2m)^2\mathcal{R}_0 \leq \mathcal{R}_{i,i}, \tag{2.18}$$

i.e. the reproduction number is higher when considering transmission within a patch as opposed to between patches.

2.3. Stochastic model

If the ODEs are not solved directly, but are instead used to define event rates for the corresponding stochastic process, then there is a distribution of possible times to invasion (t_{inv}). For sufficiently small numbers of realizations, we simulate the stochastic model using the Gillespie algorithm [39,40], but use the standard 'tau-leaping' adaptive time-step algorithm when large numbers of simulations are required [41,42].

We define the time between the first appearance of one infection in the first population ($I_1 = 1$, $t = 0$), and the first appearance of one infection in the second population ($I_2 = 1$, $t > 0$), to be the ***time to invasion***, t_{inv}. Since the ordinary differential equations (ODEs) in equation (2.10) have a unique solution associated with any given initial state, there is exactly one value of t_{inv} associated with each parameter set ($\{\beta, \gamma, N_1, N_2, m_1, m_2\}$) in the deterministic limit. Stochastic realizations yield a distribution of t_{inv}. In figure 1, we show a single stochastic realization of the model, and the corresponding time to invasion t_{inv} in that instance.

2.4. Notation summary

Our notation for variables and parameters, and the initial conditions used in all simulations and analyses, are summarized in tables 1, 2 and 3. All our simulations were performed with equal populations in the two patches ($N_1 = N_2$), as well as symmetric coupling, coupling (only one coupling parameter $m = m_1 = m_2$). We note that equal population sizes and coupling symmetry are not required for our analytical results.

3. Stochastic time to invasion

The distribution of the time to invasion (t_{inv}) is shown in figure 2 for four parameter sets ($\mathcal{R}_0 = 2, 4$; $m = 0.01, 0.1$). The histograms are each based on 10 000 stochastic simulations [41]. The blue curves show an analytical approximation that we derive below in §3.1. We present numerically computed and analytically approximated

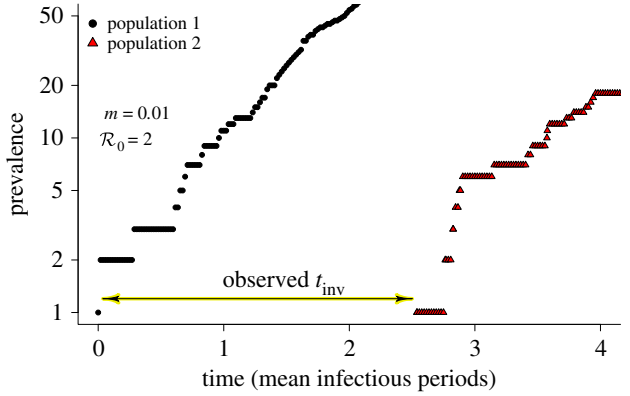

Figure 1. The *time to invasion*, t_{inv}, is the time between an initial infection in one population and the first case that appears in the other population. The figure shows a single realization of the stochastic SIR model, generated using the Gillespie algorithm [39,40] (see §2). Parameter values were $m_1 = m_2 = 0.01$, $\mathcal{R}_0 = 2$, $N_1 = N_2 = 10^5$.

Table 1. Variables of the two-population coupled SIR model.

variable	description
t	time in units of the mean infectious period, $1/\gamma$
S_1, S_2	number of susceptible individuals in each population
I_1, I_2	number of infected individuals in each population
R_1, R_2	number of removed individuals in each population

maximum-likelihood estimates (MLEs) for the coupling parameter m, given observation(s) of t_{inv}, in §§3.4 and 3.5.

3.1. Analytical approximation of time to invasion distribution

Suppose that at time $t = 0$ the system is in the initial state specified in table 3, i.e. there is a small number of individuals infected in the ***source population*** (population 1). We are interested in the time t_{inv} at which a first infection occurs in the ***receiving population*** (population 2). Until that time, there are no infections in population 2 and we will assume that t_{inv} is sufficiently short that susceptible depletion in population 1 is negligible. Thus, for $0 \leq t \leq t_{\text{inv}}$ we have $I_2(t) = 0$ and $S_1(t) \simeq N_1$, so—if we ignore demographic stochasticity[2] in population 1—equation (2.10*b*) with $i = 1$ implies that for $0 \leq t \leq t_{\text{inv}}$ we can approximate the population 1 dynamics with the single equation,

$$\frac{\mathrm{d}I_1}{\mathrm{d}t} = r_1 I_1, \tag{3.1}$$

where

$$r_1 \equiv \gamma(\mathcal{R}_{1,1} - 1), \tag{3.2}$$

and $\mathcal{R}_{i,i}$ is defined in equation (2.11*a*). Our approximation is therefore

$$I_1(t) = I_1(0)\,\mathrm{e}^{r_1 t}, \quad 0 \leq t \leq t_{\text{inv}}. \tag{3.3}$$

Given equation (3.3), and that no infections have occurred yet in population 2 (i.e. $S_2 = N_2$, $I_2 = 0$), equation (2.10*b*) with $i = 2$ specifies the (mean field[3]) rate at which infection events occur

Table 2. Parameters of the two-population coupled SIR model.

parameter	range	description
β	>0	transmission rate
$\mathcal{R}_0$	>0	basic reproduction number of the disease
γ	>0	rate of recovery from infection
m_1, m_2	$\in[0, 1]$	transmission coupling between populations
N_1	10^5	total number of individuals in population 1
N_2	10^4–10^6	total number of individuals in population 2

in population 2,

$$\mu(t) = \frac{\mathrm{d}I_2}{\mathrm{d}t} = N_2\Lambda_2 = \mu_0 e^{r_1 t}, \tag{3.4a}$$

where

$$\mu_0 = I_1(0)\,\gamma\mathcal{R}_{2,1}, \tag{3.4b}$$

and $\mathcal{R}_{2,1}$ is defined in equation (2.11*b*).[4]

In a small time interval $[t, t+\Delta t)$, we can assume that rate $\mu(t)$ is constant so the probability that an infection occurs in population 2 in this time interval is

$$\int_0^{\Delta t} \mu\,\mathrm{e}^{-\mu s}\,\mathrm{d}s = 1 - e^{-\mu\Delta t} \simeq \mu\Delta t, \tag{3.5}$$

and this is therefore also the probability that t_{inv} lies in the interval $[t, t+\Delta t)$ *given* that an infection in population 2 has not already occurred, i.e.

$$\mathrm{Prob}(t \le t_{\mathrm{inv}} < t+\Delta t \,|\, t_{\mathrm{inv}} \ge t) \simeq \mu\Delta t. \tag{3.6}$$

If we now denote the probability that invasion of population 2 occurs *before* time t by

$$F(t) = \mathrm{Prob}(0 \le t_{\mathrm{inv}} < t), \tag{3.7}$$

i.e. F is the cumulative distribution function for t_{inv}, then the probability that invasion occurs *after* time t is[5]

$$\mathrm{Prob}(t_{\mathrm{inv}} \ge t) = 1 - F(t). \tag{3.8}$$

In general, we have

$$\mathrm{Prob}(t \le t_{\mathrm{inv}} < t+\Delta t) = \mathrm{Prob}(t_{\mathrm{inv}} \ge t) \times \mathrm{Prob}(t \le t_{\mathrm{inv}} < t + \Delta t \,|\, t_{\mathrm{inv}} \ge t), \tag{3.9}$$

and hence

$$F(t+\Delta t) - F(t) \simeq \big[1 - F(t)\big]\mu(t)\Delta t. \tag{3.10}$$

Dividing by Δt and taking the limit $\Delta t \to 0$ we have

$$F'(t) = \big[1 - F(t)\big]\mu(t), \quad F(0) = 0. \tag{3.11}$$

This is a separable first-order ODE for $F(t)$, the solution of which is

$$F(t) = 1 - \exp\Big[-\int_0^t \mu(s)\,\mathrm{d}s\Big]. \tag{3.12}$$

Consequently, we can approximate the probability density function for t_{inv} by $f(t) = F'(t)$, i.e.

$$f(t) = \mu(t)\exp\Big[-\int_0^t \mu(s)\,\mathrm{d}s\Big]. \tag{3.13}$$

Inserting equation (3.4*a*) in equations (3.12) and (3.13) we obtain

$$F(t) = 1 - \exp\Big[\frac{\mu_0}{r_1}\big(1 - \mathrm{e}^{r_1 t}\big)\Big] \tag{3.14}$$

and

$$f(t) = \mu_0\,\exp\Big[r_1 t + \frac{\mu_0}{r_1}\big(1 - \mathrm{e}^{r_1 t}\big)\Big]. \tag{3.15}$$

Recall from equations (2.11), (3.2) and (3.4*b*) that r_1 and μ_0 depend implicitly on m_1 and m_2; this is important because we will need to think of f as a function of the coupling parameter(s) later.

Table 3. Initial conditions of the two-population coupled SIR model.

initial condition	value
$S_1(0)$	$N_1 - I_1(0)$
$S_2(0)$	N_2
$I_1(0)$	≥ 1
$I_2(0)$, $R_1(0)$, $R_2(0)$	0

3.2. Approximation error in time to invasion distribution

Our analysis leading to equation (3.15) was based on the approximation of pure exponential growth of cases in the first population. We can better appreciate the approximation that is being made if we recognize that the underlying process is a continuous-time branching process in the early phase during which it behaves like a simple birth–death process. During this phase, the ensemble mean number of cases in population 1 can be approximated with equation (3.3) and the associated variance is [43, p. 250]

$$\mathrm{var}[I_1](t) = I_1(0)\,\mathrm{e}^{r_1 t}(\mathrm{e}^{r_1 t} - 1). \tag{3.16}$$

To approximate the standard deviation in the force of infection from population 1 to population 2 (which we denote by σ), we scale as in equation (3.4), i.e.

$$\sigma(t) = \sigma_0\sqrt{\mathrm{e}^{r_1 t}(\mathrm{e}^{r_1 t} - 1)}, \tag{3.17a}$$

where

$$\sigma_0 = \sqrt{I_1(0)}\,\big(\gamma\mathcal{R}_{2,1}\big). \tag{3.17b}$$

We can indicate uncertainty in our analytical approximation (3.15) by replacing

$$\mu(t) \longrightarrow \mu(t) + \alpha\,\sigma(t) \tag{3.18}$$

in equation (3.13), and then, for each t, finding the maximum and minimum values of $f(t)$ for α in some specific range. Details of this calculation are given in appendix A. The thin dashed blue lines in figures 2 and 3 indicate uncertainty in $f(t)$ obtained for $\alpha \in [-0.5, 0.5]$. Note that while the dashed blue curves emphasize that the time to invasion distribution is only approximately given by the solid blue curve, they do not represent formal confidence limits; the 'α level' specified in (3.18) does not translate into a confidence limit on $f(t)$.

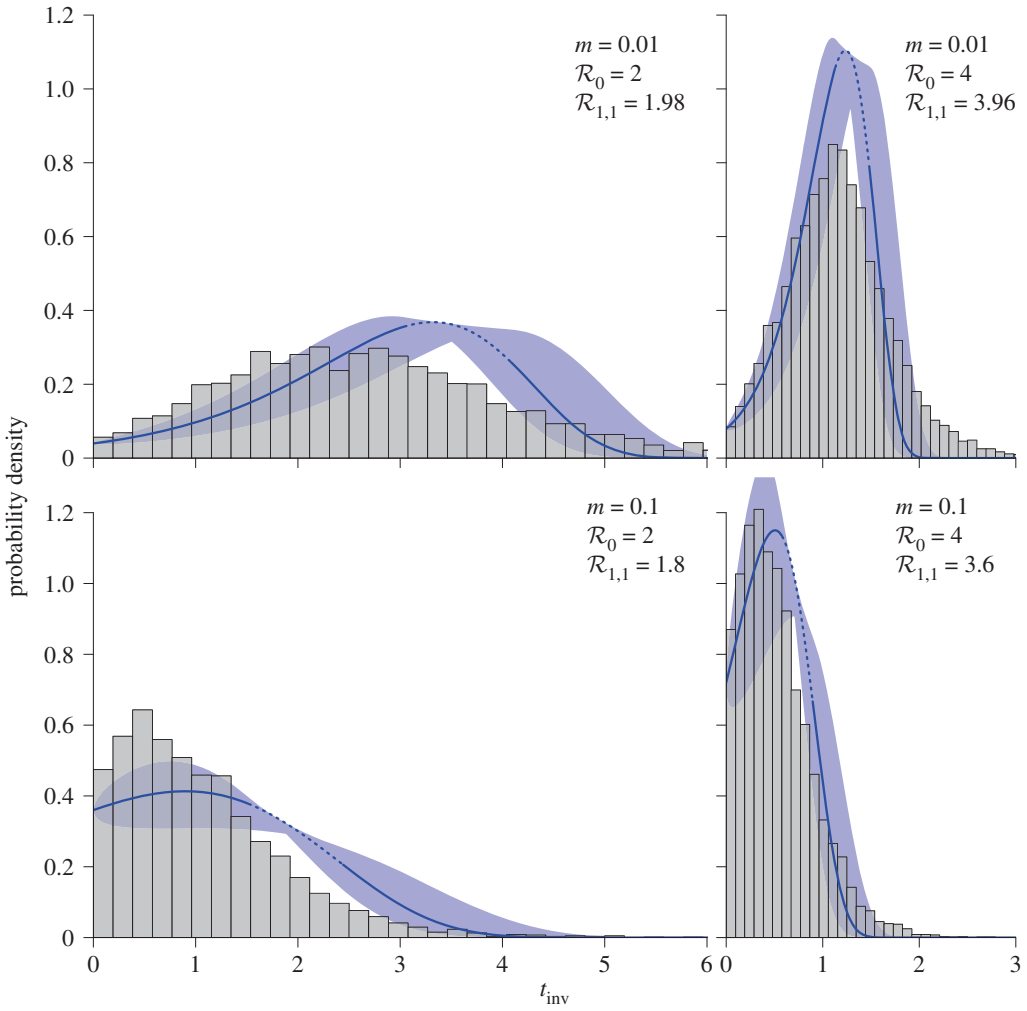

Figure 2. The probability density function for the time to invasion (t_{inv}, in units of the mean infectious period) estimated for four parameter sets ($\mathcal{R}_0 = 2, 4$; $m = 0.01, 0.1$; $N_1 = N_2 = 10^5$; $\mathcal{R}_{1,1}$ from equation (2.11)). A single infectious individual is assumed in population 1 at time 0 ($I_1(0) = 1$). Grey bars show the estimated density based on a frequency histogram constructed from 10^4 stochastic simulations (computed with the tau-leaping algorithm [41,42]) that did not fizzle (see footnotes in §3.1 and §3.3). Dark blue curves show the analytical approximation (3.15), with dotted segments corresponding to the t_{inv} range where equation (A 5) is applicable. Pale blue bands indicate uncertainty in the approximation, based on equation (A 4) with $\alpha \in [-0.5, 0.5]$ (see appendix A for derivation).

3.3. Comparison of simulations and analytical approximation

For four different parameter sets, figure 2 compares the approximate density function (3.15) with the t_{inv} distribution obtained from 10 000 realizations of the fully stochastic model.[6] As expected from the approximate formula (3.15), the probability density for t_{inv} is sensitive to both the underlying transmissibility of the pathogen ($\mathcal{R}_0$) and the degree of transmission coupling between the two patches (m).

The discrepancy between the simulations and analytical approximation in figure 2 results from variance in the epidemic curve in population 1, which is less important when the initial number of cases in population 1 is larger. To see this, note from equations (3.4) and (3.17) that the coefficient of variation in the force of infection in population 2 is

$$\frac{\sigma(t)}{\mu(t)} = \frac{\sqrt{1 - e^{-r_1 t}}}{\sqrt{I_1(0)}}, \tag{3.19}$$

which decreases rapidly with $I_1(0)$. Figure 3 shows that as $I_1(0)$ is increased, the analytical approximation of the t_{inv} distribution converges to the histogram obtained from simulations. A standard measure of the difference between two continuous probability distributions p and q is the K-L divergence [45, p. 6],

$$D_{KL}(p\|q) = \int_{-\infty}^{\infty} \big(p(x) - q(x)\big) \log \frac{p(x)}{q(x)}\, dx. \tag{3.20}$$

We define $q(x)$ to be the heights of the histogram bins, produced from stochastic simulations, in figure 3. $p(x)$ is equation (3.15) evaluated at the histogram bin midpoints. We use the K-L divergence to show the convergence of the analytical approximation of the t_{inv} probability distribution to the distribution obtained from simulations in figure 4.

3.4. Maximum-likelihood estimation of coupling parameter *m*

If we know the values of the underlying parameters ($\mathcal{R}_0$, m, N_1, N_2), then equation (3.15), or easily computable histograms like those shown in figure 2, allow us to estimate the probability of observing any particular time to invasion (t_{inv}) [46]. Our goal is to start with knowledge of

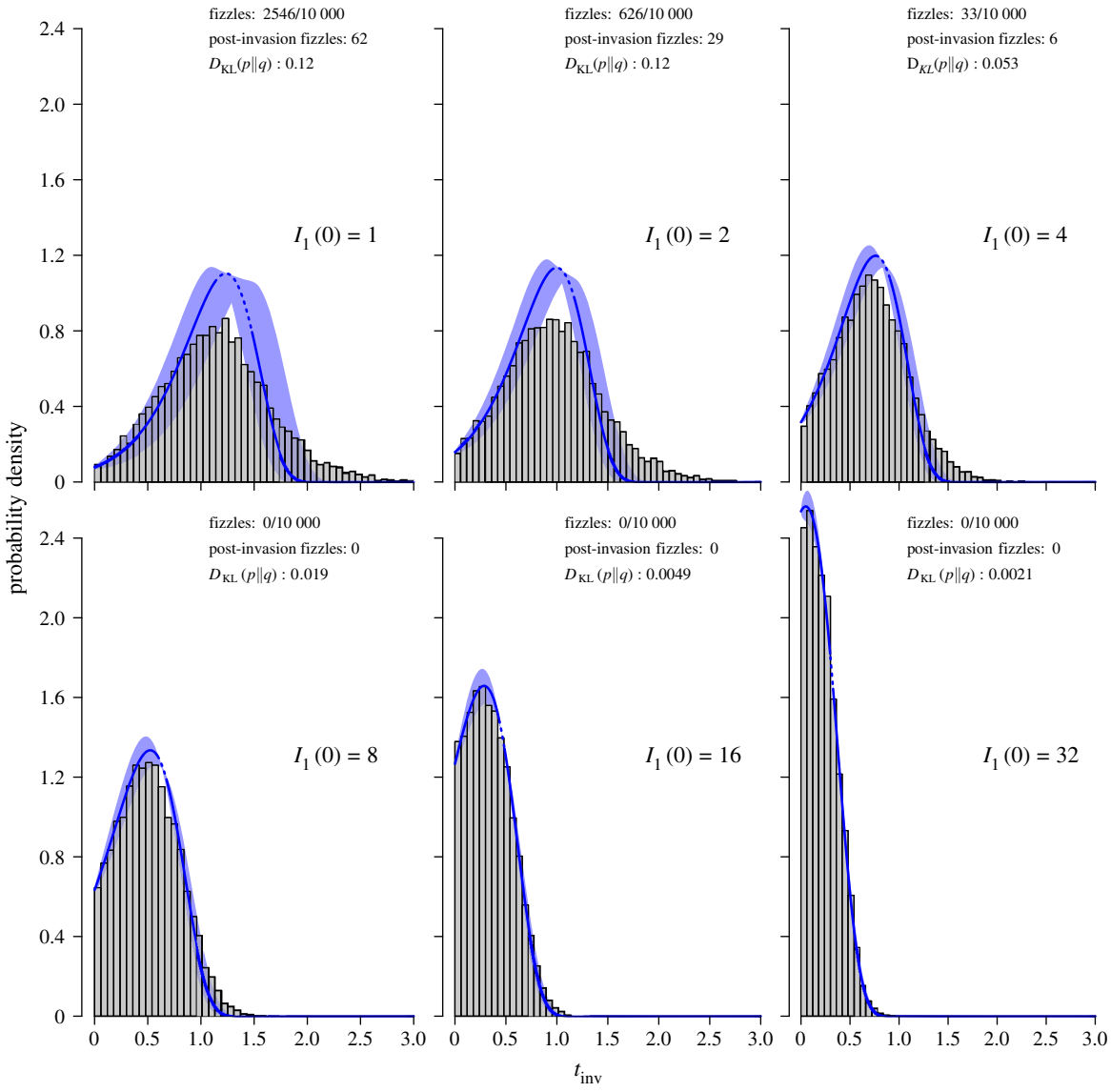

Figure 3. Probability density functions of the time to invasion t_{inv}, as in figure 2, but for a single parameter set ($\mathcal{R}_0 = 4$, $m = 0.01$, $N_1 = N_2 = 10^5$). The six panels differ in the initial numbers of infectives in population 1 ($I_1(0) \in \{1, 2, 4, 8, 16, 32\}$). Only simulations in which infection successfully spread to the second population and did not fizzle out in the first population are shown (in grey); cf. footnote in §3.3. $D_{KL}(p \,||\, q)$ refers to the Kullback–Liebler (K-L) divergence [45] (equation (3.20)), which quantifies the analytical approximation error when compared to the probability density estimated from 10^4 stochastic simulations.

— the patch population sizes (N_1, N_2),
— the disease reproduction number of the uncoupled system ($\mathcal{R}_0$),
— the mean infectious period ($1/\gamma$)

and

— one or more observations of the time to invasion (t_{inv}),

and then *estimate* the underlying transmission coupling m between the two patches. To that end, in standard fashion, we interpret the probability density of observing t_{inv} given knowledge of the underlying parameter set as the likelihood of observing m given an observation of t_{inv} [47, p. 170]. If we use our approximation (3.15), the likelihood function is

$$\mathcal{L}(m \,|\, t_{inv}) \simeq f(t_{inv}). \tag{3.21}$$

Based on this approximation, figure 5 shows the MLE of the coupling parameter m as a function of the observed time to invasion t_{inv}, for several reproduction numbers. The relationship between the MLE of m and t_{inv} is steeper for larger $\mathcal{R}_0$; consequently, to estimate m with a desired precision, more accurate estimates of t_{inv} are necessary for diseases with larger $\mathcal{R}_0$.

We can also approximate $\mathcal{L}(m \,|\, t_{inv})$ by constructing many simulation-based histograms like those in figure 2, for a range of values of m [46]. In figure 7, we show (as a heat map) a likelihood surface constructed in this way. To obtain an MLE of m for a given t_{inv} from this simulation-based likelihood surface, we (i) obtain a likelihood profile as a function of m by slicing the surface at t_{inv}, (ii) smooth the profile with a cubic spline, and then (iii) find the maximum point of the smoothed profile (figure 8).

Whether we use the analytically approximated or simulation-based likelihood, we compute confidence limits based on the likelihood ratio test (LRT) [47, ch. 6, pp. 189–194]. The LRT, applied to our MLE $\hat{m}$, assumes that the ***deviance***,

$$-2 \log \left[\frac{\mathcal{L}(m \,|\, t_{inv})}{\mathcal{L}(\hat{m} \,|\, t_{inv})} \right], \tag{3.22}$$

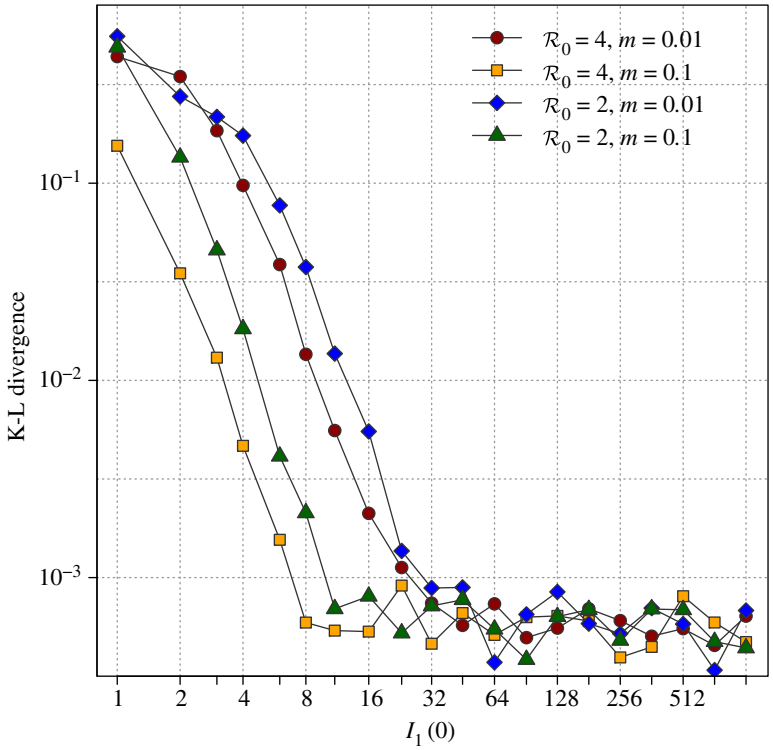

Figure 4. K-L divergence (3.20) [45] between t_{inv} distributions produced from simulations and from the analytical approximation (3.13). The K-L divergence shows the degree of difference between observed and predicted probability density distributions. Parameters used were: $\mathcal{R}_0 = 2, 4$; $m = 0.01, 0.1$; $N_1 = N_2 = 10^5$. Each point was generated by comparing the analytical approximation (3.13) with a distribution obtained from 50 000 stochastic simulations.

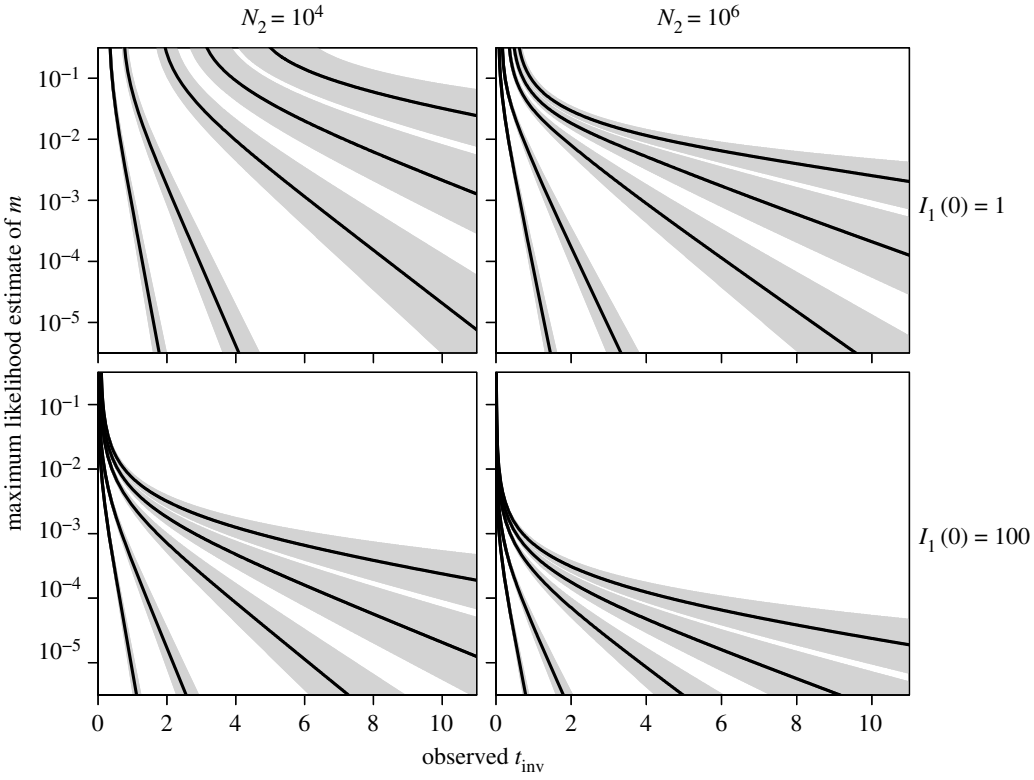

Figure 5. MLEs of coupling m versus observed time to invasion t_{inv} (in units of the mean infectious period), according to our analytical approximation (equations (3.15) and (3.21)). The size of the source population is the same in each panel ($N_1 = 10^5$). The four panels show how the relationship between the MLE of m and t_{inv} varies with the size of the receiving population ($N_2 = 10^4, 10^6$) and the initial number of infections in the source population ($I_1(0) = 1, 100$). In each panel, curves show the MLE of m for $\mathcal{R}_0 = 1.2, 1.5, 2, 4, 8$, from right to left (steeper curves correspond to larger $\mathcal{R}_0$). Grey bands under the black MLE curves indicate the effect of 10% uncertainty in the value of $\mathcal{R}_0$ (i.e. the edges of bands correspond to $0.9\mathcal{R}_0$ and $1.1\mathcal{R}_0$).

is approximately chi-squared distributed with one degree of freedom. We obtain 95% confidence limits by finding the interval along the likelihood profile of m for which the deviance is less than half the critical value of the chi-squared distribution ($\chi^2_1(0.95)/2 = 1.92$) [47, p. 192].

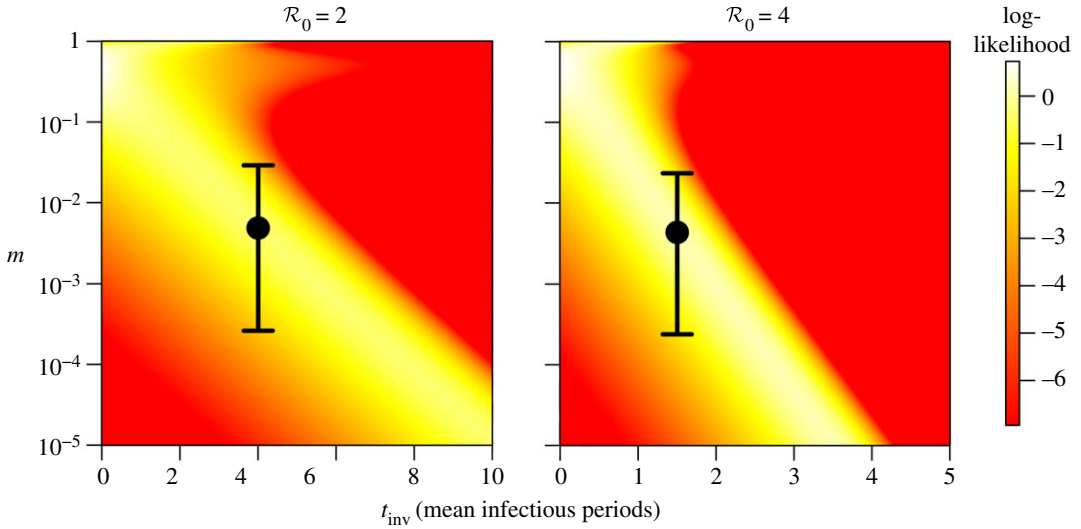

Figure 6. Likelihood of coupling parameter m given observed t_{inv}, $\mathcal{L}(m \mid t_{inv})$, computed from our analytical approximation (3.21). The fixed parameters are $N_1 = N_2 = 10^5$ and $\mathcal{R}_0 = 2$ (4) in the left (right) panel. The heavy black dot shows the MLE of m given an observed $t_{inv} = 4\ T_{inf}$ (1.5 T_{inf}), on the left (right), where T_{inf} is the mean infectious period ($T_{inf} = 1/\gamma$). Here 5% and 95% confidence limits are shown with vertical black bars. See figure 7 for a similar figure produced using stochastic simulations for comparison.

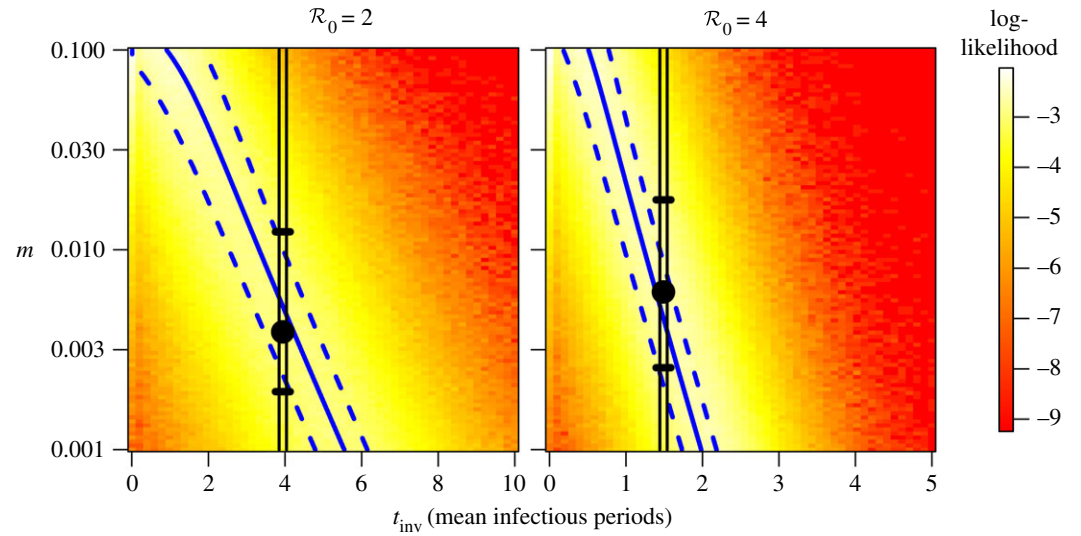

Figure 7. Likelihood of coupling parameter m given observed t_{inv}, $\mathcal{L}(m \mid t_{inv})$, computed from stochastic simulations (computed using the 'tau-leaping' algorithm [41,42]), for the same parameters as in figure 6 [$N_1 = N_2 = 10^5$, $\mathcal{R}_0 = 2$ (4)] in the left (right) panel; heavy black dot shows MLE of m given observed $t_{inv} = 4\ T_{inf}$ (1.5 T_{inf})]. The vertical black lines enclose likelihood profiles of m for the observed t_{inv}, and are shown in further detail in figure 8. Twenty-five percent and 75% confidence limits are shown with horizontal black bars. The solid blue curves in each panel show the MLE of m according to the analytical approximation (3.21) and correspond to particular curves in figure 5. The dashed blue curves show 25% and 75% confidence limits for the analytical approximation (see appendix B for details).

The MLE ($\hat{m}$) and confidence interval for m, given an observation of t_{inv}, are shown with a black dot and error bars in figures 6 and 7; figure 6 is based on our analytical approximation, while figure 7 is based on simulations (see appendix B for computational details). In figure 7, the solid blue curve shows the MLE ($\hat{m}$) as a function of t_{inv} obtained from our analytical approximation (3.21), and the dashed blue curves show the analytically approximated confidence bands. Because it was constructed using analytical formulae, the time required to create figure 6 was negligible, whereas more than 12 h of computation time were required to create figure 7.

3.5. MLE based on multiple observations of time to invasion

If multiple events of disease spread from one population to the other have been observed then much more accurate estimation of the transmission coupling parameter m is possible. It is important to emphasize in this context that since we are aiming to estimate a parameter of the social contact network—as opposed to a disease parameter—there is no need to restrict attention to repeated invasions by a single pathogen. Independent invasions by unrelated infectious diseases with the same mode of transmission could, in principle, be

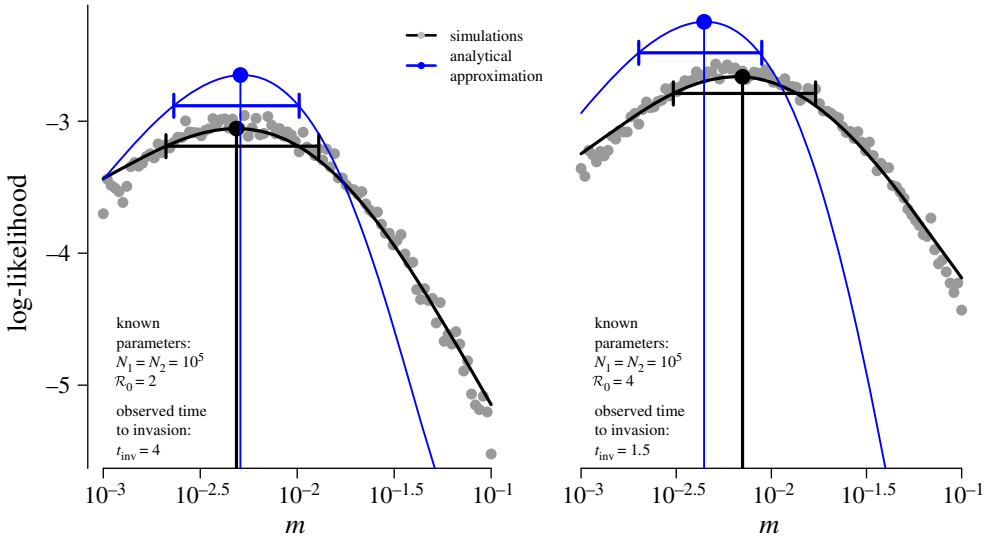

Figure 8. Likelihood profiles for the coupling parameter m. Black curves show the likelihood profile obtained from stochastic simulations (cf. figure 7) and blue curves are obtained from our analytical approximation (3.21). Heavy dots show the MLE and error bars show the 25% and 75% confidence limits. The grey dots correspond to the column enclosed with vertical black lines in the heat map in figure 7; we smooth these log-likelihood values with a cubic spline and define the MLE and confidence limits using the spline.

just as valuable for this purpose. Estimates of m from independent invasions would require the assumption that m does not change between events, along with accurate estimates of disease parameters, $\mathcal{R}_0$ and γ, for each invading disease.

Suppose n independent invasions have been observed and let θ_i denote the set of observations $\{\mathcal{R}_0, \gamma^{-1}, t_{\text{inv}}\}$ associated with the ith invasion event. Then the likelihood of the coupling parameter being m, given this sequence of n observed invasions, is

$$\mathcal{L}(m\,|\{\theta_1, \ldots, \theta_n\}) = \prod_{i=1}^{n} \mathcal{L}(m\,|\,\theta_i). \tag{3.23}$$

Each factor $\mathcal{L}(m\,|\,\theta_i)$ can be approximated using equation (3.15) or via a simulation-based, smoothed likelihood profile, as in figure 8.

Figure 9 shows four examples of how an estimate of m using the simulation-based approach improves as the number of observed invasions increases from 1 to 64. In each of four panels, the 64 invasions are assumed to be by the same disease (so the same $\mathcal{R}_0$ and mean infectious period). Exactly how the MLE and 95% confidence intervals change as additional invasions are observed depends on the sequence in which the observations occur. Each panel of figure 9 shows three extreme cases, in which the 64 t_{inv} observations occur from (i) shortest to longest, (ii) longest to shortest, and (iii) from the median of the 64 observations to median of the remaining 63, and so on. The equivalent figure based on the analytical approximation (3.21) is shown in figure 10.

4. Discussion

We have explored the feasibility of using the time taken for an infectious disease to spread from one population to another (the time to invasion, t_{inv}) to estimate the degree of social contact between two populations. We quantified the degree of social contact with the proportion (m) of time that individuals typically spend outside their home region.

We have considered only the most idealized situation in which there are only two populations and the basic reproduction number, $\mathcal{R}_0$, and mean infectious period, $1/\gamma$, of the disease are known precisely. Even so—if based on a single observed disease invasion—the confidence intervals we obtain for the degree of coupling (m) stretch over an order of magnitude (figure 8), which therefore provides only crude information about the social connectivity of the two populations. However, if multiple invasions are observed, much more accurate estimation of m is possible (figure 9), and the independent invasions need not be of the same disease (§3.5).

We estimated the likelihood profile for the coupling parameter m in two ways (figure 8), one based on large numbers of stochastic simulations and the other based on an analytical approximation that we derived in §3.1. The simulation approach is more accurate (figures 2 and 9 versus 10), but significantly so only if the number of cases in the source population is very small when the estimate is made (figure 3). The large computational expense of the simulation approach could be reduced by, for example, iterated filtering [48] beginning from the analytically derived MLE, but simulations would be hard to justify if $\gtrsim 10$ cases had already occurred in the source population (figure 3).

Our analytical approximation facilitates exploration of how the relationship between observed t_{inv} and MLE of m depends on underlying disease characteristics—such as $\mathcal{R}_0$ and the mean infectious period—and on uncertainty in estimates of those properties (figure 5).

It is worth noting that our analysis leading to an analytical approximation of the likelihood function (equation (3.21)) depends only on the epidemic growing exponentially in the source population, not on any details of the natural history of infection. For more complex models (e.g. models that include non-exponential waiting time distributions [11,12,49,50]), equation (3.2) for the exponential growth rate r_1 in the source population would simply be replaced by a different expression for the more realistic model in question.

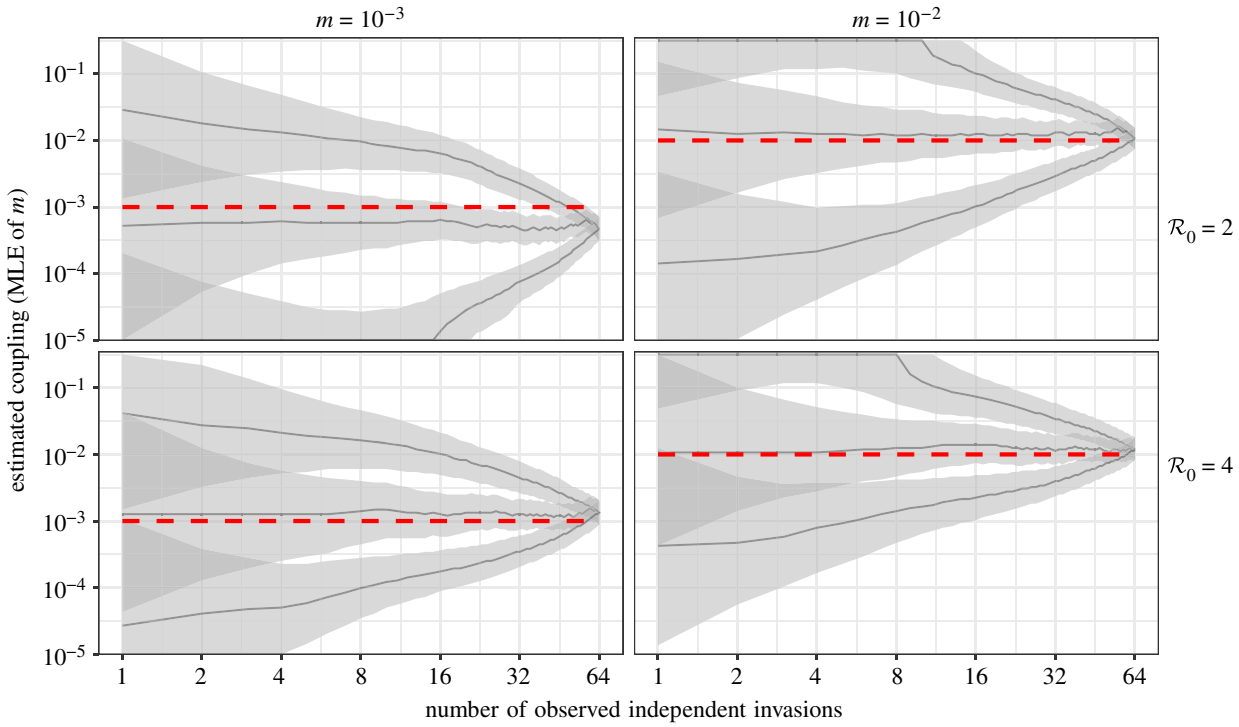

Figure 9. Estimates of the coupling parameter (m) improve as more independent invasion events are observed. The underlying $\mathcal{R}_0$ and coupling (m) are indicated above and to the right of the panels, and the underlying m is shown with a red dashed line. Populations sizes are $N_1 = N_2 = 10^5$ in all panels. In each case, 64 invasion events were simulated with the stochastic model (§2.3). The lower and upper curves show the MLE of m estimated from the subset of the 64 simulations corresponding to the largest and smallest observed times to invasion (note that high observed t_{inv} implies low coupling m, and vice versa). The MLEs shown with the middle curve correspond to the subset of simulations for which the observed t_{inv} was closest to the median. The shaded regions shows 95% confidence limits. In this figure we show estimation of coupling m using stochastic simulations (cf. figures 7 and 8, and §3.5). See figure 10 for the equivalent graphs based on the analytical approximation (3.21).

4.1. Limitations

If attempts are made to apply our methodology to real epidemics, a number of limitations are important to bear in mind.

— The time to invasion t_{inv} can be difficult to estimate because of incomplete or inaccurate reporting, reporting delays, asymptomatic cases, and lack of temporal resolution in reporting (especially for historical data).
— If multiple invasions are observed, with long breaks between them, the possibility of changes in population characteristics in the times between epidemics should be considered. This can be a particularly significant concern when examining historical epidemics separated by decades or centuries.
— In general, changes in human behaviour and other factors may alter the social contact network *during* an epidemic and consequently the coupling of subpopulations of a meta-population.

4.2. Possible further developments

There are several natural directions for enhancement of the methods developed in this paper.

— The exponential growth approximation in §3.1 could be replaced with a better approximation of an epidemic curve. For example, instead of equation (3.3), a logistic curve could be used to approximate initial growth in the source population beyond the exponential phase [51,52], or an approximate or exact analytical solution to a transmission model [53–55] could be used estimate the entire epidemic curve in the source population. Any phenomenological or mechanistic model could be fitted to an observed source epidemic, and could potentially lead to better analytical approximations of $\mu(t)$, the expected rate at which new infections occur in the receiving population. This estimate would replace equation (3.4*a*) and, after insertion in equation (3.13), lead to an alternative version of equation (3.15) for the probability density of the time to invasion. Such analyses would potentially be valuable in situations involving extremely weak coupling, where the probability of invasion after the exponential phase might be substantial.
— Rather than relying on an approximation to the epidemic curve in the source population, the actual time series of observed cases could be used instead of equation (3.3) (for example, by assuming each case is infectious for exactly the mean infectious period). This would lead to a (presumably more accurate) estimate of the time to invasion distribution. This approach would preclude fully analytical investigations, but would presumably provide more accurate results (in weak coupling situations).
— In a meta-population with more than two populations, the time at which a first case occurs in each subpopulation could be used to inform the coupling in the system. In principle, it could turn out to be easier to estimate the *average* inter-population transmission coupling when there are more subpopulations. On the

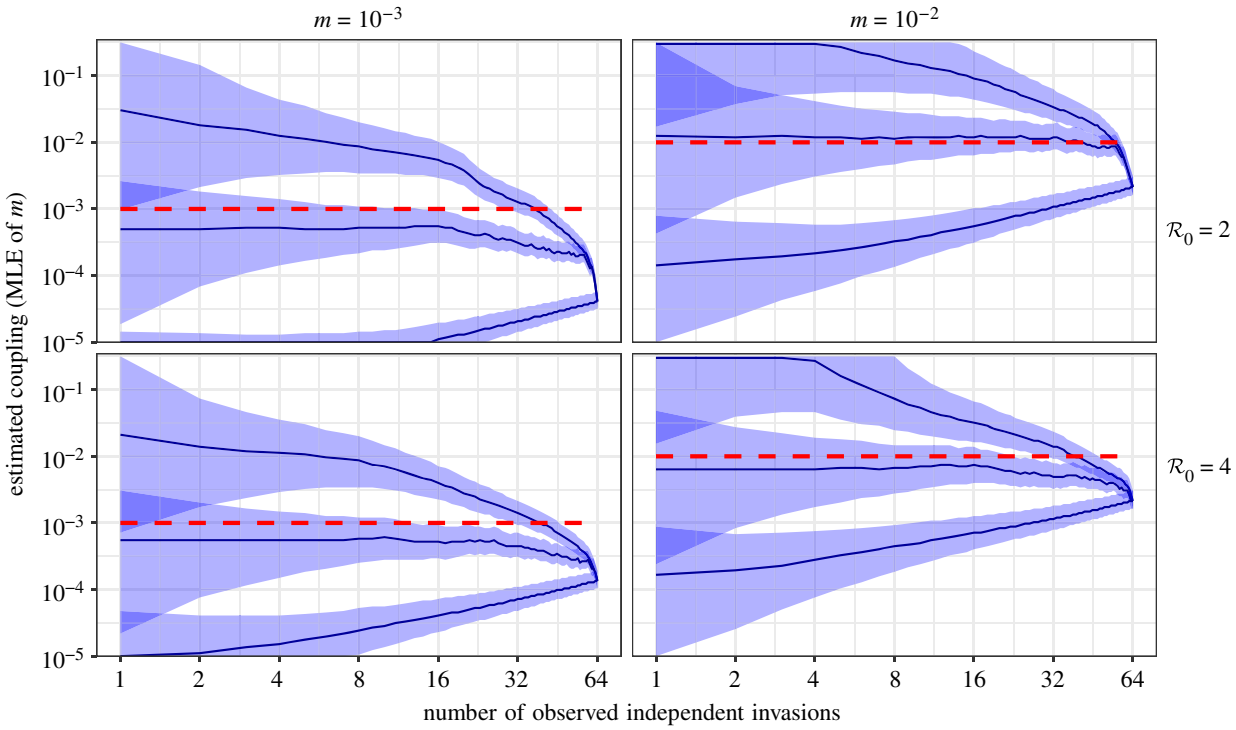

Figure 10. The equivalent of figure 9 based on the analytical approximation (3.15) rather than simulations.

other hand, potentially different degrees of coupling between each pair of subpopulations increases the range of possible contact networks. For familiar, circulating pathogens, existing knowledge of disease parameters and population sizes can be used to estimate coupling for the typically smaller populations that are below the critical community size [56] and therefore experience local extinctions and re-invasions. For newly invading pathogens, coupling throughout the meta-population can be estimated, but will be limited by the accuracy of disease parameter estimates during the invasion.

— While our analytical results apply to any 2×2 coupling matrix, all of our illustrative examples assumed equal coupling ($m_1 = m_2$). A systematic analysis of the effects of unequal coupling parameters would be valuable. This seems especially relevant in the context of coupling between large cities and small towns, since one would expect a larger proportion of residents of a small town to visit a large city regularly than vice versa.

— In figure 5, we indicated the effect of uncertainty in $\mathcal{R}_0$. A more systematic and complete analysis of the effects of uncertainty in estimates of non-coupling parameters would be valuable.

— We have focused on the time to invasion, but if there are more than two subpopulations then the locations of the source subpopulations that seed each invasion could also be used to constrain estimates of connectivity.

— If age-stratified incidence or mortality data are available, more detail about transmission coupling could be extracted, in principle. Different age-groups have been observed to make contact at different rates [57], and the age distribution of infections in the source population along with the age of the first case in the receiving population could better inform estimations of inter-population coupling than the time to invasion alone.

— In some situations, information about travel volumes and destinations may be available, in which case ways to use such data to constrain connectivity estimates (such as with the use of Bayesian priors [58]) could be useful.

— In a situation where multiple independent invasions can be observed, an estimate of the coupling parameter m from earlier events, along with another estimate of m from later events, might have non-overlapping confidence intervals. This would be evidence of changes in the underlying social contact network.

Our analysis in this paper has shown that while estimating coupling from the time to invasion is difficult, it is possible. Enhancing methods of doing so will advance understanding of the mechanisms and predictability of infectious disease outbreaks in meta-populations.

Data accessibility. This article has no additional data.

Authors' contributions. K.H. carried out project conceptualization, development of theoretical results, software programming and simulation, writing and editing of manuscript text, figures and tables. D.E. also carried out project conceptualization, development of theoretical results, writing and editing of manuscript.

Competing interests. We declare we have no competing interests.

Funding. K.H. was supported by an Ontario Graduate Scholarship (OGS). D.E. was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Acknowledgements. We are grateful for helpful comments from Ben Bolker, Jonathan Dusho and Alun Lloyd.

Endnotes

[1]Similar formulations of cross-coupling can be found in the literature, such as Murray & Cliff [32], Lloyd & May [33], Lloyd & Jansen [34]. We derive our formulation of coupling on a meta-population fully in [35, §4.3.2], which we omit here since we are dealing with only two populations. An alternative approach to the two-patch invasion problem, based on a model with explicit travel between patches, is presented by Yan *et al.* [36].

[2]In the stochastic setting, with probability $(1/\mathcal{R}_{1,1})^{I_1(0)}$, an outbreak in population 1 fizzles out without causing a full blown epidemic [43, §7.6.2, p. 321]. Nevertheless, the second population is sometimes infected before the outbreak fizzles out in the first population. This effect is larger for lower $\mathcal{R}_0$, and for sufficiently small $\mathcal{R}_0$ must be taken into account to understand the expected distribution of t_{inv}. We ignore fizzles in our analysis, but in figures 2 and 3 we indicate the number of simulations that fizzled and were therefore ignored.

[3]The *mean field* refers to the ensemble mean of all stochastic realizations.

[4]In the derivation that follows, we assume that the incidence in population 1 must be approximated in order to estimate the distribution of the time to invasion, t_{inv}. However, if the actual trajectory of incidence in population 1 is known, then this distribution can be computed exactly, since the force of infection on population 2 can be calculated at each point in time.

[5]The derivation presented here follows along the lines of standard survival analysis, where our hazard function is characterized by the force of infection on population 2 by population 1 (e.g. [44, p. 13]).

[6]We keep a stochastic simulation only if two conditions are satisfied: (i) the second population is eventually infected ($I_2(t) > 0$ for some $t > 0$), and (ii) the first population does not fizzle. We consider the outbreak to have fizzled in population 1 if the prevalence in that population drops to zero before the cumulative proportion of the population infected reaches the level corresponding to the peak of the deterministic epidemic curve (at the peak, the proportion susceptible, $S_1(t)/N_1$, is $1/\mathcal{R}_{1,1}$ (see equation (2.11))). The number of susceptibles in the first population, $S_1(t)$, does not increase, and decreases as individuals become infected. Once the condition $S_1(t)/N_1 < 1/\mathcal{R}_{1,1}$ is satisfied, $\mathrm{d}I_1/\mathrm{d}t$ (equation (2.10b)) remains strictly negative for all future times. Thus, the condition defining a non-fizzle is $I_1(t) > 0$ for all $t > 0$ until $S_1(t)/N_1 < 1/\mathcal{R}_{1,1}$.

Appendix A. Approximation error on t_{inv} distribution

The ensemble mean and variance of the force of infection from the source population (1) to the receiving population (2) are given in equations (3.4) and (3.17), respectively. To quantify uncertainty on the distribution of the time to invasion of population 2, we must evaluate the integral in equation (3.13) for $\mu(t) + \alpha\,\sigma(t)$ rather than $\mu(t)$, i.e. we must calculate

$$f_\alpha(t) = \left[\mu(t) + \alpha\,\sigma(t)\right] \exp\left\{ -\int_0^t \left[\mu(s) + \alpha\,\sigma(s)\right] \mathrm{d}s \right\}. \tag{A 1}$$

(Note that $f(t)$ in equation (3.13) corresponds to $f_0(t)$ in this notation.) To evaluate the integral in equation (A 1) explicitly, we use

$$\int_0^t \sqrt{\mathrm{e}^{rs}(\mathrm{e}^{rs} - 1)}\,\mathrm{d}s = \frac{1}{r}\left[\sqrt{\mathrm{e}^{rt}(\mathrm{e}^{rt} - 1)} - \log\left(\sqrt{\mathrm{e}^{rt} - 1} + \sqrt{\mathrm{e}^{rt}}\right)\right]. \tag{A 2}$$

Thus, with μ and σ given by equations (3.4) and (3.17), respectively, and writing r for r_1 to reduce clutter, we obtain the explicit expression,

$$f_\alpha(t) = \left[\mu_0 \mathrm{e}^{rt} + \alpha\,\sigma_0\sqrt{\mathrm{e}^{rt}(\mathrm{e}^{rt} - 1)}\right] \times \exp\left\{\frac{\mu_0}{r}\left(1 - \mathrm{e}^{rt}\right)\right\} \times \exp\left\{ -\alpha\frac{\sigma_0}{r}\left[\sqrt{\mathrm{e}^{rt}(\mathrm{e}^{rt} - 1)} - \log\left(\sqrt{\mathrm{e}^{rt} - 1} + \sqrt{\mathrm{e}^{rt}}\right)\right]\right\}. \tag{A 3}$$

For a given α range ($\alpha_{\min} \le \alpha \le \alpha_{\max}$, where normally $\alpha_{\min} = -\alpha_{\max}$), we then define upper and lower error estimates,

$$f_{\text{U}}(t) = \max_\alpha \left\{ f_\alpha(t) : \alpha_{\min} \le \alpha \le \alpha_{\max} \right\} \tag{A 4a}$$

and

$$f_{\text{L}}(t) = \min_\alpha \left\{ f_\alpha(t) : \alpha_{\min} \le \alpha \le \alpha_{\max} \right\}, \tag{A 4b}$$

which correspond to the blue bands in figures 2 and 3. For any given t, at least one of the upper and lower estimates is obtained at an edge of the α range; solving $\partial f_\alpha/\partial\alpha = 0$ for α, we find a single critical point

$$\alpha_{\text{crit}}(t) = \frac{\sqrt{\mathrm{e}^{rt} - 1}(r - \mu_0\,\mathrm{e}^{rt}) + \mu_0\,\mathrm{e}^{rt/2}\log\left(\sqrt{\mathrm{e}^{rt} - 1} + \mathrm{e}^{rt/2}\right)}{\sigma_0\left[\mathrm{e}^{rt/2}(\mathrm{e}^{rt} - 1) - \sqrt{\mathrm{e}^{rt} - 1}\,\log\left(\sqrt{\mathrm{e}^{rt} - 1} + \mathrm{e}^{rt/2}\right)\right]}. \tag{A 5}$$

Appendix B. Numerical details of simulation-based likelihood

This appendix relates to the construction of figure 7, as described in §3.4.

For each of 100 m values, we measured time to invasion t_{inv} from 10^4 stochastic simulations using the adaptivetau package in R [41], and grouped these t_{inv} values into 100 bins on the t_{inv} axis. More precisely, our 100 m values, which we refer to as m_i, were spaced logarithmically between 0.001 and 0.1. For each m_i, and for $\mathcal{R}_0 = 2, 4$, we produced $n_{\text{sim}} = 10^4$ simulations and measured the corresponding t_{inv} for each simulation. We then divided the full range of resulting t_{inv} values into 100 bins, b_j. We produced a grid where Cell(i, j) contained the number of simulations with $m = m_i$ and t_{inv} in bin b_j. We used the grid of m vs. t_{inv} simulation frequencies to produce likelihoods of t_{inv} given m,

$$\mathcal{L}(t_{\text{inv}}|m_i) \approx \frac{\text{Cell}(i, j)}{n_{\text{sim}}}. \tag{B 1}$$

We produced a full grid of log-likelihoods, i.e. $\log\mathcal{L}(t_{\text{inv}}|m_i)$ (figure 7). We select the bin b_j that contains the observed t_{inv}. The log-likelihoods of column j yield the likelihood profile of the observed t_{inv} with respect to m, and the cell with the maximum likelihood indicates the MLE of m given t_{inv} (figure 8).

References

1. Anderson RM, May RM. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
2. Hethcote HW. 2000 The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653. (doi:10.1137/S0036144500371907)
3. Earn DJD, Rohani P, Bolker BM, Grenfell BT. 2000 A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670. (doi:10.1126/science.287.5453.667)
4. Bauch CT, Earn DJD. 2003 Transients and attractors in epidemics. *Proc. R. Soc. Lond. B* **270**, 1573–1578. (doi:10.1098/rspb.2003.2410)
5. Earn DJD. 2009 Mathematical epidemiology of infectious diseases. In *Mathematical biology* (eds MA Lewis, MAJ Chaplain, JP Keener, PK Maini), vol. 14 of IAS/ Park City Mathematics Series, pp. 151–186. Providence, RI: American Mathematical Society.
6. London W, Yorke JA. 1973 Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *Am. J. Epidemiol.* **98**, 453–468. (doi:10.1093/oxfordjournals.aje.a121575)
7. Yorke JA, London W. 1973 Recurrent outbreaks of measles, chickenpox and mumps. II. Systematic

differences in contact rates and stochastic effects. *Am. J. Epidemiol.* **98**, 468–482. (doi:10.1093/oxfordjournals.aje.a121576)

8. He D, Earn DJD. 2007 Epidemiological effects of seasonal oscillations in birth rates. *Theor. Popul. Biol.* **72**, 274–291. (doi:10.1016/j.tpb.2007.04.004)
9. Anderson D, Watson R. 1980 On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67**, 191–198. (doi:10.1093/biomet/67.1.191)
10. Feng ZL, Thieme HR. 2000 Endemic models with arbitrarily distributed periods of infection I: fundamental properties of the model. *SIAM J. Appl. Math.* **61**, 803–833. (doi:10.1137/S0036139998347834)
11. Lloyd AL. 2001 Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc. R. Soc. Lond. B* **268**, 985–993. (doi:10.1098/rspb.2001.1599)
12. Wearing HJ, Rohani P, Keeling MJ. 2005 Appropriate models for the management of infectious diseases. *PLoS Med.* **2**, 621–627.
13. Nishiura H, Eichner M. 2007 Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiol. Infect.* **135**, 1145–1150. (doi:10.1017/S0950268806007618)
14. Conlan AJK, Rohani P, Lloyd AL, Keeling M, Grenfell BT. 2010 Resolving the impact of waiting time distributions on the persistence of measles. *J. R. Soc. Interface* **7**, 623–640. (doi:10.1098/rsif.2009.0284)
15. Newman MEJ. 2002 Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 1–11.
16. Newman MEJ. 2010 *Networks: an introduction*. New York, NY: Oxford University Press.
17. Watts DJ, Strogatz S. 1998 Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)
18. Arino J. 2003 A multi-city epidemic model. *Math. Popul. Stud.* **10**, 175–193. (doi:10.1080/08898480306720)
19. Grenfell BT, Bjornstad ON, Kappey J. 2001 Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723. (doi:10.1038/414716a)
20. Viboud C, Bjornstad ON, Smith DM, Simonsen L, Miller MA, Grenfell BT. 2006 Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451. (doi:10.1126/science.1125237)
21. Sattenspiel L. 2009 *The geographic spread of infectious diseases: models and applications*. Princeton Series in Theoretical and Computational Biology. Princeton, NJ: Princeton University Press.
22. May RM, Anderson RM. 1984 Spatial heterogeneity and the design of immunization programs. *Math. Biosci.* **72**, 83–111. (doi:10.1016/0025-5564(84)90063-4)
23. Sattenspiel L. 1987 Population structure and the spread of disease. *Hum. Biol.* **59**, 411–438.
24. Cliff AD, Haggett P. 1988 *Atlas of disease distributions: analytic approaches to epidemiologic data*. Oxford, UK: Basil Blackwell.
25. Cliff AD, Haggett P, Smallman-Raynor M. 1993 *Measles: an historical geography of a major human viral disease, from global expansion to local retreat, 1840–1990*. Oxford, UK: Blackwell Publishers.
26. Sattenspiel L, Dietz K. 1995 A structured epidemic model incorporating geographic-mobility among regions. *Math. Biosci.* **128**, 71–91. (doi:10.1016/0025-5564(94)00068-B)
27. Ferguson NM, May RM, Anderson RM. 1997 Measles: persistence and synchronicity in disease dynamics. In *Spatial ecology* (eds D Tilman, P Kareiva), vol. 30 of Monographs in Population Biology, pp. 137–157. Princeton, NJ: Princeton University Press.
28. Earn DJD, Rohani P, Grenfell BT. 1998 Persistence, chaos and synchrony in ecology and epidemiology. *Proc. R. Soc. Lond. B* **265**, 7–10. (doi:10.1098/rspb.1998.0256)
29. Rohani P, Earn DJD, Grenfell BT. 1999 Opposite patterns of synchrony in sympatric disease metapopulations. *Science* **286**, 968–971. (doi:10.1126/science.286.5441.968)
30. Earn DJD, Levin SA, Rohani P. 2000 Coherence and conservation. *Science* **290**, 1360–1364. (doi:10.1126/science.290.5495.1360)
31. Watts DJ, Muhamad R, Medina DC, Dodds PS. 2005 Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci. USA* **102**, 11 157–11 162. (doi:10.1073/pnas.0501226102)
32. Murray G, Cliff AD. 1977 A stochastic model for measles epidemics in a multi-region setting. *Trans. Inst. Br. Geogr.* **2**, 158–174. (doi:10.2307/621855)
33. Lloyd AL, May RM. 1996 Spatial heterogeneity in epidemic models. *J. Theor. Biol.* **179**, 1–11. (doi:10.1006/jtbi.1996.0042)
34. Lloyd AL, Jansen VAA. 2004 Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. *Math. Biosci.* **188**, 1–16. (doi:10.1016/j.mbs.2003.09.003)
35. Hempel K. 2018 Detecting epidemic coupling among geographically separated populations. PhD thesis, McMaster University, Hamilton, Ontario.
36. Yan AW, Black AJ, McCaw JM, Rebuli N, Ross JV, Swan AJ, Hickson RI. 2018 The distribution of the time taken for an epidemic to spread between two communities. *Math. Biosci.* **303**, 139–147. (doi:10.1016/j.mbs.2018.07.004)
37. Diekmann O, Heesterbeek JAP, Metz JAJ. 1990 On the definition and the computation of the basic reproduction ratio RO in models for infectious-diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382. (doi:10.1007/BF00178324)
38. van den Driessche P, Watmough J. 2002 Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180**(Sp. Iss.), 29–48. (doi:10.1016/S0025-5564(02)00108-6)
39. Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434. (doi:10.1016/0021-9991(76)90041-3)
40. Gillespie DT. 1977 Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.* **81**, 2340–2361. (doi:10.1021/j100540a008)
41. Johnson P. 2019 adaptivetau: tau-leaping stochastic simulation. R package version 2.2–3. See https://CRAN.R-project.org/package=adaptivetau.
42. Gillespie DT. 2007 Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55. (doi:10.1146/annurev.physchem.58.032806.104637)
43. Allen LJS. 2010 *An introduction to stochastic processes with applications to biology*, 2nd edn. Upper Saddle River, NJ: Pearson Education Inc.
44. Cox DR, Oakes D. 1984 *Analysis of survival data*. London, UK: Chapman & Hall.
45. Kullback S. 1997 *Information theory and statistics*. Mineola, NY: Courier Corporation.
46. Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A. 2011 Statistical inference for stochastic simulation models—theory and application. *Ecol. Lett.* **14**, 816–827. (doi:10.1111/j.1461-0248.2011.01640.x)
47. Bolker BM. 2008 *Ecological models and data in R*. Princeton, NJ: Princeton University Press.
48. Ionides EL, Breto C, King AA. 2006 Inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18 438–18 443. (doi:10.1073/pnas.0603181103)
49. Krylova O, Earn DJD. 2013 Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J. R. Soc. Interface* **10**, 20130098. (doi:10.1098/rsif.2013.0098)
50. Champredon D, Dushoff J, Earn DJD. 2018 Equivalence of the Erlang SEIR epidemic model and the renewal equation. *SIAM J. Appl. Math.* **78**, 3258–3278. (doi:10.1137/18M1186411)
51. Ma J, Dushoff J, Bolker BM, Earn DJD. 2014 Estimating initial epidemic growth rates. *Bull. Math. Biol.* **76**, 245–260. (doi:10.1007/s11538-013-9918-2)
52. Earn DJD, Ma J, Poinar H, Dushoff J, Bolker BM. 2020 Acceleration of plague outbreaks in the second pandemic. *Proc. Natl Acad. Sci. USA* **117**, 27 703–27 711. (doi:10.1073/pnas.2004904117)
53. Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)
54. Wilson EB, Worcester J. 1945 The law of mass action in epidemiology. *Proc. Natl Acad. Sci. USA* **31**, 24–34. (doi:10.1073/pnas.31.1.24)
55. Harko T, Lobo FS, Mak M. 2014 Exact analytical solutions of the susceptible–infected–recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* **236**, 184–194.
56. Bartlett MS. 1957 Measles periodicity and community size. *J. R. Stat. Soc. Ser. A* **120**, 48–70. (doi:10.2307/2342553)
57. Mossong J *et al.* 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, 381–391. (doi:10.1371/journal.pmed.0050074)
58. Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB. 2013 *Bayesian data analysis*. Boca Raton, FL: Chapman and Hall/CRC.