

Impact of the Infection Period Distribution on the Epidemic Spread in a Metapopulation Model

Elisabeta Vergu^{1*}, Henri Busson^{1,2,3}, Pauline Ezanno^{2,3}

1 UR341 Mathématiques et Informatique Appliquées, INRA, Jouy-en-Josas, France, **2** UMR1300 Bio-agression, Epidémiologie et Analyse de Risques, INRA, Nantes, France, **3** UMR1300 Bio-agression, Epidémiologie et Analyse de Risques, ENVN, Nantes, France

Abstract

Epidemic models usually rely on the assumption of exponentially distributed sojourn times in infectious states. This is sometimes an acceptable approximation, but it is generally not realistic and it may influence the epidemic dynamics as it has already been shown in one population. Here, we explore the consequences of choosing constant or gamma-distributed infectious periods in a metapopulation context. For two coupled populations, we show that the probability of generating no secondary infections is the largest for most parameter values if the infectious period follows an exponential distribution, and we identify special cases where, inversely, the infection is more prone to extinction in early phases for constant infection durations. The impact of the infection duration distribution on the epidemic dynamics of many connected populations is studied by simulation and sensitivity analysis, taking into account the potential interactions with other factors. The analysis based on the average nonextinct epidemic trajectories shows that their sensitivity to the assumption on the infectious period distribution mostly depends on R_0 , the mean infection duration and the network structure. This study shows that the effect of assuming exponential distribution for infection periods instead of more realistic distributions varies with respect to the output of interest and to other factors. Ultimately it highlights the risk of misleading recommendations based on modelling results when models including exponential infection durations are used for practical purposes.

Citation: Vergu E, Busson H, Ezanno P (2010) Impact of the Infection Period Distribution on the Epidemic Spread in a Metapopulation Model. PLoS ONE 5(2): e9371. doi:10.1371/journal.pone.0009371

Editor: Nick Monk, University of Nottingham, United Kingdom

Received: May 29, 2009; **Accepted:** January 29, 2010; **Published:** February 26, 2010

Copyright: © 2010 Vergu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was carried out with the financial support of the ANR (French National Research Agency, <http://www.agence-nationale-recherche.fr>) under the Programme Agriculture et Développement Durable, project ANR-05-PADD-014. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: elisabeta.vergu@jouy.inra.fr

Introduction

The use of mathematical models in the study of epidemic dynamics, their mechanisms and their prevention and control provided valuable insights and contributed to draw up the global picture of epidemics occurrence, patterns and management [1]. Therefore, considering the impact of mathematical models in epidemiology, it is important to build them on realistic assumptions.

Most often, in mathematical models for epidemic spread, sojourn times in infectious states are exponentially distributed, which means that the probability of recovery per unit of time is constant, regardless of the time elapsed since infection. This classical assumption is routinely used for mathematical tractability-based reasons. Although this is sometimes an acceptable approximation, it is not realistic in general.

Several papers have explored the impact of more realistic non-exponentially distributed disease stages on the epidemic dynamics (summarized by various criteria, such as basic reproductive number (R_0), extinction probability, speed of propagation, epidemic burden, intensity of epidemic peak, etc) in single populations [2–5].

Most of these studies were motivated by the epidemiology of childhood viral diseases, such as measles, in order to reproduce their observed persistence patterns exhibiting localized extinctions. Since models making the assumption of exponential distributed infectious times were not able to mimic these observed patterns, Keeling and Grenfell [3,6] tested constant sojourn times and provided

simulations characterized by increasing persistence and therefore closer to observations. Independently, theoretical arguments were provided by Lloyd [7,8] supporting the idea that constant infectious periods had a destabilizing effect on the global dynamics leading more often to extinction. This noticeable contradiction was discussed by Keeling and Grenfell [9] who proposed as a possible explanation the fact that parameters have to be updated with respect to the model, when the same observed phenomenon was described by different models. These authors argued that the results provided by theoretical studies were not realistic since in these works parameters were not updated when changing the model.

Recent papers in theoretical epidemiology have explored the question of R_0 estimation from data with respect to other modelling assumptions, especially in the early phase of an epidemic [4,10–13]. Their authors emphasized that caution has to be paid when calculating R_0 from estimations of the initial growth rate, since the equation relating these two parameters varies with respect to the distribution of the infection period.

Another category of modelling studies [14–19] have focused on epidemic dynamics in a metapopulation context, an important framework to explore when attempting to understand epidemic dynamics at a large scale [20–24]. Indeed, human or animal populations are not isolated; they influence each other (by exchanging individuals, for instance). Therefore, pathogens spread is the result of the complex interplay between intra-population events and inter-population interactions.

Despite the existence of numerous studies on these topics, less attention has been paid to connect both aspects, non-exponentially distributed infectious periods and metapopulation global dynamics. The metapopulation context can be illustrated, for instance, by the transmission of an animal infection disease in a group of connected farms. Since models are often used to test the effectiveness of various interventions implemented at the scale of a region including many herds and since the assumption on the infectious stage distribution could affect model outcomes, it is important to correctly assess its effect.

Here, we focus on the analytical and computer-based exploration of different consequences of the introduction of non exponentially distributed sojourn times in epidemic models developed for metapopulation contexts. As already noticed, this is a question of non negligible importance, especially when models are built to serve as predictive tools. Indeed, the consequences of non realistic or inadequate modelling choices could provide biased results and hence orient towards inappropriate recommendations. Besides metapopulation aspects, our study differs from recent work focused on R_0 estimation as far as we adopt a different point of view on data. Whereas these papers [10,12,13] are interested in the estimation of R_0 from data in the early stage of an outbreak, when little information about a disease is available, and focus on the relationship between R_0 and the observed growth rate, we address situations where some knowledge about average epidemiological parameters of a disease is available and when the modelling is used to predict propagation at different time horizons. Indeed, such a situation can occur when information such as mean and range of infection duration are available based on expert opinion, but few data was collected. As an example, it was shown that for the contagious bovine pleuropneumonia, a respiratory disease of cattle exhibiting very diverse clinical patterns, constant infection durations were more appropriate than exponentially distributed ones for describing experimental data [25].

In this study, we explore possible discrepancies in forecasted dynamics in relation to model assumptions. More precisely, we are interested in: (i) what would happen (in terms of global criteria such as global epidemic burden, epidemic duration, extinction) if the infectious period was constant or gamma instead of exponentially distributed in a metapopulation context, when the epidemic is described by a *SIR* (Susceptible-Infectious-Removed) model in each patch and infection spreads between patches as a consequence of individuals movements; (ii) how the differences between these global criteria calculated for couples of distributions of the infectious period would vary with respect to R_0 , the transmission rate, the mean infection duration, the intensity of flows between patches and the network structure in a stochastic metapopulation model (consisting of $p \times (SIR)$ local models). Our approach is decomposed in two steps. First we focus on analytical developments of a specific criterion, the probability of early extinction, in single and two coupled populations. Analytical calculation is dedicated to the impact of the distribution of infectious period on the probability of no secondary infections and of extinction after g generations by extending the work of Keeling and Grenfell [3] to the case of gamma distribution in single populations. The two-population case is also explored concerning the probability that an infectious individual will cause no secondary cases. Second, since analytical explorations are intractable for more than two populations, we tackle the p -population case, where populations are connected through various networks and with different coupling intensities, by simulations using event-driven and individual-based approaches and statistical analysis of simulated data. We enlarge the framework by focusing on other aspects of an epidemic (such as epidemic size and duration, etc) in addition to the probability of early extinction. A sensitivity analysis (conducted through an analysis of variance) of various epidemic

outputs (expressed as differences between outputs under different assumptions on the infectious period distribution) with respect to input factors such as R_0 and network topology is performed.

Our paper is structured in several parts as follows: we first present the basic mathematical formulation of epidemic models we use, the main concepts necessary to the analytical explorations and the main lines of the sensitivity analysis. In the theoretical part of the results, analytical expressions of probabilities of early extinction and of extinction after g generations for single populations and two coupled populations are derived. In the applied subsection of results, computer-based explorations are performed for studying the sensitivity of the effects of infection period distribution to the input factors in a metapopulation framework. A general discussion is provided in the last section.

Throughout the article terms *patch* and *population* are used as synonyms to designate a local community with homogeneous contacts.

Methods

State variables and transition probabilities of stochastic models are first described. Then, we define probabilities of interest in single and two-population models. Finally, we provide details on simulations performed for models including more than two populations and on statistical analysis of simulated data. Important parameters, variables and functions used throughout the paper are defined in Table 1.

Formulation of the Stochastic Model

Analytical calculations and simulations are performed in a continuous-time stochastic framework. Mathematical models used here are classical epidemic models including three distinct disease states and assuming a density-dependent force of infection. The demography is not included, a situation that could fit to rapid infections, spreading and developing on short time scales. The metapopulation context is taken into account by the inclusion of movements between local populations (into and from all compartments) which are considered to have time invariant sizes on average.

Counts of individuals in each patch and for each disease state are represented by random variables defined on a discrete state space. The state of the global system at time t in patch i is represented by the vector $\mathbf{X}_i(t) = (S_i(t), I_i(t), R_i(t))$, where $S_i(t)$, $I_i(t)$ and $R_i(t)$ represent the number of susceptible, infectious and recovered individuals respectively.

For the Markovian case, where the probability of future behaviours of the process depends only on the present state, all sojourn times are exponentially distributed. This implies that, for instance, the rate of recovery is constant with respect to time (illustrating the memorylessness of exponential distribution). Transition probabilities for elementary changes in random variables in each patch i are defined as follows:

$$\begin{aligned}
 \mathbb{P}[S_i(t + \Delta t) = S_i(t) - 1, I_i(t + \Delta t) \\
 &= I_i(t) + 1 | \mathbf{X}(t)] = \beta_i S_i I_i \Delta t / N_i + \mathbf{o}(\Delta t) \\
 \mathbb{P}[S_i(t + \Delta t) = S_i(t) - 1, S_j(t + \Delta t) \\
 &= S_j(t) + 1 | \mathbf{X}(t)] = \sigma_{ij} S_i \Delta t + \mathbf{o}(\Delta t) \\
 \mathbb{P}[I_i(t + \Delta t) = I_i(t) - 1, R_i(t + \Delta t) \\
 &= R_i(t) + 1 | \mathbf{X}(t)] = \gamma I_i \Delta t + \mathbf{o}(\Delta t) \\
 \mathbb{P}[I_i(t + \Delta t) = I_i(t) - 1, J_j(t + \Delta t) \\
 &= J_j(t) + 1 | \mathbf{X}(t)] = \sigma_{ij} I_i \Delta t + \mathbf{o}(\Delta t) \\
 \mathbb{P}[R_i(t + \Delta t) = R_i(t) - 1, R_j(t + \Delta t) \\
 &= R_j(t) + 1 | \mathbf{X}(t)] = \sigma_{ij} R_i \Delta t + \mathbf{o}(\Delta t)
 \end{aligned} \tag{1}$$

Table 1. Summary of important parameters, variables and functions.

Name	Expression	Definition
$S(t)$		Number of susceptibles at t
$I(t)$		Number of infecteds at t
$R(t)$		Number of recovereds at t
T_I (day)		R.v.* infectious period
γ (day ⁻¹)		Recovery rate
t_I (day)	$1/\gamma$	Mean of T_I
β (day ⁻¹)		Transmission rate
σ (day ⁻¹)		Intensity of migration
λ	$\beta S/N$	Rate of generation of new cases
R_0	β/γ	Basic reproductive number
R	$\beta S/N\gamma$	Effective reproductive number
X_r		R.v.* number of secondary cases generated by an infectious individual
$f_n(u) \equiv \Gamma(n, n\gamma)$	$\frac{(\gamma n)^n u^{n-2} e^{-u\gamma n}}{(n-1)!}$	Probability density function of T_I
T_1 (day)		Mixed r.v.* time spent in the first population by an individual during his infectious period in a 2-population model
$g(u, \tau)$	cf. eqs. (4) and (14)	Continuous component of T_1
m_u	$e^{-\sigma_1 u}$	Mass component of T_1
$P_n(r)$	cf. eqs. (3) and (9)	Probability of $X_r = r$ in single populations
$P_n^{2pop}(r)$	cf. eq. (4)	Probability of $X_r = r$ in a 2-population model
$P_n(0)$	cf. eqs. (5) and (9)	Probability of $X_r = 0$ in single populations
$P_n^{2pop}(0)$	cf. eq. (6)	Probability of $X_r = 0$ in a 2-population model
$P_{n,ext}(g)$	cf. eq. (7)	Extinction probability after g generations of infecteds in single populations
$P_n^{2pop}(0 1 \text{ jump})$	cf. eq. (15)	Probability of $X_r = 0$ in a 2-population model given that the infectious individual generating secondary cases moves only once between populations
$G_n(z)$	cf. eq. (10)	Probability generating function of X_r
C_f^Y	cf. eq. (8)	Sensitivity index equal to the contribution of factor f to the variation in Y

*R.v. = random variable.

All the variables or functions indexed by n are related to the distribution $\Gamma(n, n\gamma)$ of T_I .

doi:10.1371/journal.pone.0009371.t001

where γ is the recovery rate, β_i and N_i are respectively the transmission rate and the population size of patch i and $\sigma_{ij} = \sigma_{ji}$ (with $\sigma_{ii} = 0$) is the per capita rate of migration from patch i to patch j .

When the sojourn time in infectious compartment is not exponentially distributed, the associated jump process is no longer Markovian. An appropriate way to deal with this case is to consider the explicit history of each individual rather than a population-based perspective, since in the absence of the lack-of-

memory property, removal times of individuals now depend on the time of entering the infective state. In an individual-based formulation, each individual ind of a population i can experience one of the three possible types of transitions during the interval $(t, t + \Delta t)$:

$P(ind \text{ gets infected in } (t, t + \Delta t) | ind \text{ was susceptible at } t)$

$$= \beta_i I_i \Delta t / N_i$$

$P(ind \text{ moves from patch } i \text{ to patch } j \text{ in } (t, t + \Delta t) | ind \text{ was in patch } i \text{ at } t)$

$$= \sigma_{ij} \Delta t \tag{2}$$

$P(ind \text{ recovers in } (t, t + \Delta t) | ind \text{ has been infectious for } d_{ind} \text{ time at } t)$

$$= \frac{f_n(d_{ind}) \Delta t}{1 - F_n(d_{ind})}$$

where f_n and F_n are respectively the probability density and distribution functions of the random variable T_I which represents the sojourn time in the infectious state. We take $f_n \equiv \Gamma(n, n\gamma)$ (with integer $n \geq 1$) since this formulation embraces the three distributions of T_I under study: exponential (for $n = 1$), constant (when $n \rightarrow \infty$) and gamma (for $1 < n < \infty$). This last case could correspond to a situation where the infectious state of a disease consists in n distinct stages, which can be differentiated based on distinct symptoms. Indeed, the duration of a process decomposed in a sequence of n independent stages can be modelled by a $\Gamma(n, n\gamma)$ distribution (which can be written as the sum of n independent exponentially distributed random variables of parameter γ). This restriction to integer values for the shape parameter n (which corresponds to the Erlang distribution) is very useful in modelling since it captures real scenarios. The mean of the infectious period is the same for all distributions, $E(T_I) = 1/\gamma = t_I \forall n$.

Definition of Probabilities of No Secondary Cases in Single and Two Coupled Populations and of Extinction after g Generations of Infecteds in Single Populations

The probabilities of producing no secondary cases related to early extinction and of extinction after g generations summarize the main interesting dynamical behaviours, since they provide information on extinction at different stages in the epidemic evolution. They also allow the comparison between models incorporating gamma distributed infectious sojourn times (with exponential distributed and constant infectious periods are the two extremes) with equal means and with different variances. Prior to performing the comparison of these probabilities provided in the results section, we define them in the context of single and two-coupled populations.

Similarly to Keeling and Grenfell [3], the individual level perspective is considered by introducing the random variable X_r representing the number of secondary cases generated by an infectious individual (R_0 is then the expected value of X_r). Subscripts indicating the population (as in eq. (1)) are not used in order to avoid overloading notations. According to [3], the probability of generating r secondary cases by any infectious individual, depending on the shape parameter n of f_n , is given by the expression:

$$P_n(r) = \mathbb{P}(X_r = r) = \int_0^\infty f_n(u) e^{-\beta S u / N} \frac{(\beta S u / N)^r}{r!} du, \tag{3}$$

if we assume that new infections are realisations of a homogeneous

Poisson process and are produced at a constant rate $\lambda = \beta S/N$. That is, the proportion S/N of susceptible individuals remains constant at the beginning or for the duration of an epidemic, which could be generally not so unrealistic for large populations.

In the case of two coupled populations, the production of new cases in each patch is assumed to follow homogeneous Poisson processes of intensities λ_1 and λ_2 respectively. Let us also define T_1 , the random variable representing the time spent by an individual in the first patch (assumed to be his origin) during his infectious period u , over all successive sojourns. Given $T_I = u$, T_1 is mixed, since it consists of a mass at u (associated with the Dirac mass equal to 1 in u and to 0 elsewhere) and an absolutely continuous component on $(0, u)$. The mass component, noted by m_u , corresponds to the case where the individual does not leave his initial patch. The continuous component, noted by $g(u, \tau)$, is a continuous function of variables u (the infectious duration of the individual) and τ (the part of u spent in population 1). We also introduce r_1 , the variable representing the number of secondary cases that this individual generates in the first patch, given that he globally produces r secondary infections. Then, the probability of generating r secondary cases in two coupled populations by any infectious individual moving from one population to the other, depending on n and noted by $P_n^{2pop}(r)$ is defined by:

$$P_n^{2pop}(r) = \mathbb{P}(X_r = r \text{ secondary cases in the two patches}) = \int_0^\infty f_n(u) \left(\sum_{r_1=0}^r \int_0^{u^-} g(u, \tau) e^{-\lambda_1 \tau} \frac{(\lambda_1 \tau)^{r_1}}{r_1!} e^{-\lambda_2 (u-\tau)} \frac{(\lambda_2 (u-\tau))^{r-r_1}}{(r-r_1)!} d\tau + m_u e^{-\lambda_1 u} \frac{(\lambda_1 u)^r}{r!} \right) du. \tag{4}$$

From eqs. (3) and (4), probabilities of generating no secondary cases ($r=0$), are obtained for single and two coupled populations:

$$P_n(0) = \int_0^\infty f_n(u) e^{-\lambda u} du \tag{5}$$

and

$$P_n^{2pop}(0) = \int_0^\infty f_n(u) e^{-\lambda_1 u} \left(\int_0^{u^-} g(u, \tau) e^{(u-\tau)(\lambda_1 - \lambda_2)} d\tau + m_u \right) du, \tag{6}$$

respectively, where $m_u = e^{-\sigma_1 u}$ and $g(u, \tau)$ has to be calculated.

Another criterion for comparing the impact of f_n shape on epidemic dynamics is the extinction probability after g generations, if the production of new cases is interpreted as a branching process. Similarly to [3], this probability can be expressed in a single population under a recursive form:

$$P_{n,ext}(g) = \sum_{r=0}^\infty P_n(r) P_{n,ext}(g-1)^r. \tag{7}$$

Ingredients for a Sensitivity Analysis of the Outputs of Epidemic Dynamics in Metapopulations

Analytical investigations become less tractable for more than two populations, especially for situations where the connections between populations are not homogeneous, i.e. not all populations are connected or migration intensities are different between couples of populations. Moreover, the matter of extinction in metapopulations was explored by simulations elsewhere [1].

In order to explore more broadly the interplay between infectious period shape and other factors (such as R_0 and network connectivity) in terms of impacts on metapopulation dynamics, a sensitivity analysis evaluating how model outputs vary when entries are modified is performed. Special attention is paid to the potential influence of the network (mainly in terms of mean connectivity) through which populations are connected.

The sensitivity analysis is conducted through an analysis of variance (ANOVA). In this kind of analysis, the variance of each *dependent variable* is partitioned into components due to different *factors* in order to see if the variability observed in the dependent variable is due to variations in factors or results from ‘‘by-chance’’ effects. *Factors* are independent variables whose values are controlled and varied by the experimenter. *Dependent variables* represent the response that is observed as a consequence of the independent variables being manipulated.

In the following, details are given on the input *factors*, the *dependent variables* and the formulas used to assess the effect of each factor.

Several input *factors* are tested: (i) the shape of the infectious period distribution, (ii) R_0 , (iii) the transmission rate (β), (iv) the mean infection duration (t_I), (v) the migration rate (σ) and (vi) the network topology.

Three special shapes for the infectious duration distribution belonging to gamma family are considered: $T_I \sim \Gamma(1, \gamma)$ ($\equiv \text{Exp}(\gamma)$; exponential), $T_I = 1/\gamma$ (constant) and $T_I \sim \Gamma(3, 3\gamma)$ (gamma).

For each of the remaining input factors, several values (reported in Table 2) are chosen to illustrate a large panel of possible realistic situations in terms of capacity of transmission, average duration of infection as well as coupling and connectivity between populations. In these scenarios, R_0 (calculated as $\beta * t_I$) lies on a range from 0.4 to 9.

Several networks structures are proposed in order to test the influence of the connectivity structure, both in terms of mean connectivity and degree distributions (since not all nodes in a network have the same number of connections). These two characteristics are known to be particularly important for mechanisms by which diseases spread over networks [26]. First, the networks tested here are chosen to represent a large panel of topologies according to the two mentioned features. They are the ring, the star, the completely connected network, the homoge-

Table 2. Values taken by input factors in ANOVA.

Parameters	Definition	Values
Distribution	Infectious period distribution	$T_I \sim \text{Exp}(\gamma)$, $T_I \sim \Gamma(3, 3\gamma)$, $T_I = 1/\gamma$
β (day ⁻¹)	Transmission rate	0.08, 0.1, 0.2, 0.3
t_I (day)	Mean infection duration	5, 10, 15, 30
σ (day ⁻¹)	Intensity of migration	10^{-4} , 2.5×10^{-4} , 7.5×10^{-4} , 10^{-3}
Network	Network topology	Complete, Ring, Star, Homogeneous-random (6, 10, 20, 30, 40, 50)*, Scale-free (6, 10, 20, 30, 40, 50)*

All the values for the 4-uples (β , t_I , γ and Network) were considered for each of the three shapes of the infection duration distribution (960 different scenarios). In ANOVA all these factors were treated as ordinal variables by transforming numerical values in ordered categories.

*Numbers in brackets represent mean connectivities of networks that were tested. doi:10.1371/journal.pone.0009371.t002

neous-random network and the scale-free network. For our metapopulation comprising 100 population (or nodes) the mean connectivity is equal to 99 for the completely connected network, 2 for the ring network and 1.98 for the star network. For the homogenous-random and scale free networks several average degrees of connectivity were tested: 6, 10, 20, 30, 40 and 50. Second, for a given mean connectivity, the networks tested are also different according to the distribution of their degree of connectivity: e.g. a homogenous-random network is characterized by a Poisson distribution, whereas a scale-free network has a power-law tail and comprises both highly connected hubs and very low connected nodes [26].

All the scenarios obtained by crossing the values given in Table 2 are simulated: 960 scenarios are tested for each of the three distributions of the infectious period. For each scenario, the parameters are taken equal for all populations.

The following *dependent variables* (represented here by *epidemic outputs*) are considered for each simulated scenario corresponding to a fixed set of parameters and many simulation runs: (i) the proportion of minor epidemics, (ii) the epidemic duration, (iii) the date of intra-population epidemic peak averaged over all the populations of the network, (iv) the date and (v) the size of the epidemic peak of the metapopulation (considered as a whole) and (vi) the final epidemic size. All but (i) criterion are calculated for major epidemics only. Minor epidemics (assimilated to the probability of early extinction) and major epidemics (i.e. epidemics that do not undergo early extinction) are distinguished according to the proportion of population that is finally infected. More precisely, for each scenario $P(\text{minor epidemic}) = P(\text{early extinction}) = P(s(\infty) > 0.95)$ (where $s(\infty)$ represents the proportion of susceptibles in the total population at the end of the outbreak) and is calculated as $\frac{\sum_{i=1}^{nruns} 1_{s(\infty) > 0.95}}{nruns}$ (where $nruns$ is the number of simulation runs by scenario). Similarly, major epidemics correspond to epidemic trajectories which satisfy the constraint $s(\infty) \leq 0.95$.

Criteria (iv), (v) and (vi) are calculated at both individual and population levels. For each scenario and for all but (i) dependent variables, means and variances over major epidemics among simulation runs are calculated.

Since we focus our interest on the shape of the distribution of the infectious period, we separately consider this factor. Therefore, for each scenario, three variants are considered for each dependent variable except for the proportion of minor epidemics: 'E-G' (difference between the value of the output simulated with the exponential distributed infectious period and the value corresponding to the gamma distributed infectious period), 'G-C' (difference between the value of the output simulated with the gamma distributed infectious period and the value corresponding to a constant infectious period) and 'E-C' (difference between the value of the output simulated with the exponential distributed infectious period and the value corresponding to a constant infectious period).

The global contribution of *factor f* to the variations in the *dependent variable Y* is assessed through a sensitivity coefficient, C_f^Y , including the principal effect and first-order interactions in which factor *f* is involved [27]:

$$C_f^Y = \frac{SS_f^Y + \frac{1}{2} \sum_{l,l \neq f} SS_{f,l}^Y}{SS_{tot}^Y} \tag{8}$$

In eq. (8) $SS_{tot}^Y = \sum_f (SS_f^Y + 1/2 \sum_{l,l \neq f} SS_{f,l}^Y) + SS_{error}^Y = SS_{reg}^Y + SS_{error}^Y$ is the total sum of squares (i.e. the sum of squared

distances from any point in the dataset to the mean of the data) for the dependent variable *Y*, SS_f^Y is the sum of squares related to the principal effect of factor *f* on *Y* and $SS_{f,l}^Y$ represents the sum of squares related to the effect of interactions between factors *f* and *l* on *Y*. In our analysis, one ANOVA is carried out separately for every dependent variable *Y* which corresponds to each of the three variants ('E-G', 'G-C' and 'E-C') of epidemic outputs previously described. *f* corresponds to each of our input factors. Logarithmic transformation is used for the dependent variables. This or other appropriate transformations of dependent variables are sometimes needed to render linear their relationships with factors, in order to fulfil one of the main assumptions of standard ANOVA (which assumes a linear regression).

In order to more specifically identify factors influencing the proportion of minor epidemics (accounting for early extinction) we also calculate partial rank correlation coefficients (PRCC) for this variable. PRCCs allow the quantification of the non linear association between a given input and the dependent variable when controlling for the other input parameters.

Simulations are performed using a model including $p=100$ populations of $N_i=20$ individuals each, coupled by identical migration rates ($\sigma_{ij} = \sigma_{ji} = \sigma, \forall i,j$) and with equal transmission rates ($\beta_i = \beta, \forall i$). The initial condition is generated by randomly seeding 5 of 100 populations of the network ($I_i(0) = 1$ and $S_i(0) = 19$ for 5 of 100 populations and 0 and 20 respectively for all remaining ones; $R_i(0) = 0, \forall i$). For each scenario, 300 simulation runs are performed using a time continuous event-driven approach based on eqs. (1) and (2) and including three categories of possible events: (i) infection of a susceptible individual (to which corresponds an

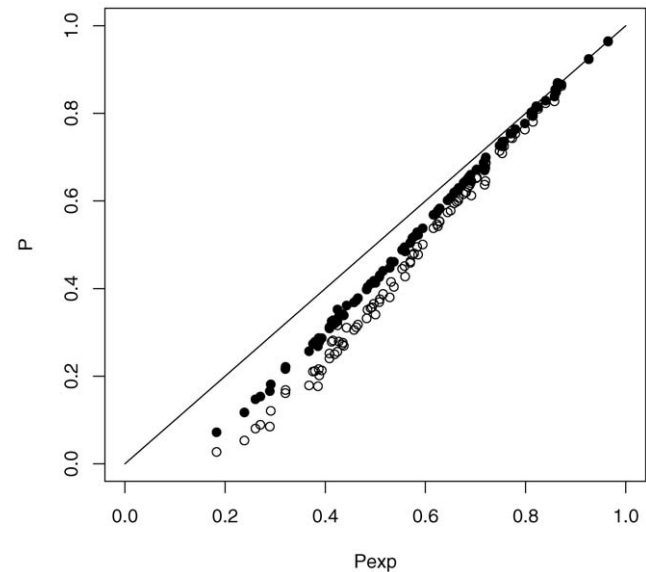


Figure 1. Comparison of probabilities of no secondary cases in a two-population model. On y-axis *P* refers to $P_3^{2pop}(0)$ (closed circles) and $P_2^{2pop}(0)$ (open circles). On x-axis *Pexp* refers to $P_1^{2pop}(0)$. For each of the three probabilities, 100 points with different parameter combinations were generated ($\lambda_1, \lambda_2, \sigma_1$ and σ_2 were drawn from exponential distributions and γ was taken equal to 1). The points represent means over 100000 Monte-Carlo simulations of a time continuous event-driven approach: one infectious individual is introduced in one population and the probability of no secondary cases is calculated based on the time spent in each population. Estimated average standard deviation for computed values was below 0.001. All parameters and variables are explained in Table 1. doi:10.1371/journal.pone.0009371.g001

exponential distributed sojourn time in compartment S of parameter $\beta_i S_i I_i / N_i$, (ii) recovery of an infectious individual (to which corresponds exponential, gamma distributed or constant sojourn times in compartment I of parameters γ , n and $n\gamma$ and $1/\gamma$ respectively) and (iii) migration of an individual between populations regardless of his disease state (to which corresponds an exponential distributed sojourn time in the population of origin, i , before leaving for population j , of parameter σ_{ij}). In the case where all transitions are Markovian we use the classical Gillespie's algorithm (details in [1], p201) that simulates the time until the next event and its type. For gamma distributed or constant infectious duration, when the transitions from infectious to recovered states are no longer Markovian, it is necessary to explicitly simulate the history of each individual instead of using the Gillespie's algorithm. In this individual-based approach, the time until the next event considering the metapopulation as a whole will be the minimum over all the individual times. Algorithms are implemented in C language. The various network topologies underlying the metapopulation are generated using *sispread* software [28].

Results

Comparison of Probabilities of No Secondary Cases and of Extinction after g Generations of Infecteds in Single Populations with Respect to the Infectious Period Distribution

First, we expanded the work of Keeling and Grenfell [3] who treated the case of constant distributions for one population, to the gamma distributed sojourn times. We then extended our analysis to the case of two populations.

For a gamma distributed infectious period $T_I \sim \Gamma(n, n\gamma)$ (i.e. $f_n(u) = (\gamma n)^n u^{n-1} e^{-u\gamma n} / (n-1)!$, where n takes positive nonzero integer values), eq. (3) becomes:

$$P_n(r) = \int_0^\infty \frac{(\lambda u)^r (\gamma n)^n u^{n-1} e^{-u(\gamma n + \lambda)}}{r!(n-1)!} dt \tag{9}$$

$$= \left(\frac{R/n}{1+R/n} \right)^r \left(\frac{1}{1+R/n} \right)^n \Gamma_n^r,$$

where $R = R_0 S / N = \beta S / N \gamma = \lambda / \gamma$ is the effective reproductive number and $\Gamma_n^r = \binom{n+r-1}{r} = (n+r-1)! / ((n-1)! r!)$. This generalizes the expressions found in [3] for the cases where $T_I \sim \text{Exp}(\gamma)$ and $T_I = 1/\gamma$ which can be recalculated from eq. (9) by making $n=1$ and $n \rightarrow \infty$, as $P_1(r) = \left(\frac{R}{1+R} \right)^r \left(\frac{1}{1+R} \right)$ and $P_\infty(r) = e^{-R} \frac{R^r}{r!}$ respectively.

The probability of generating no secondary infections, $P_n(0)$, defined in eq. (5), can be viewed as a particular case of the probability generating function (p.g.f.) of X_r . Let $G_n(z) = \sum_{r=0}^\infty P_n(r) z^r$ ($z \in (0,1)$) denote the p.g.f. of X_r when $T_I \sim \Gamma(n, n\gamma)$. Using eq. (9) $G_n(z)$ becomes:

$$G_n(z) = \left(\frac{1}{1+(R(1-z))/n} \right)^n. \tag{10}$$

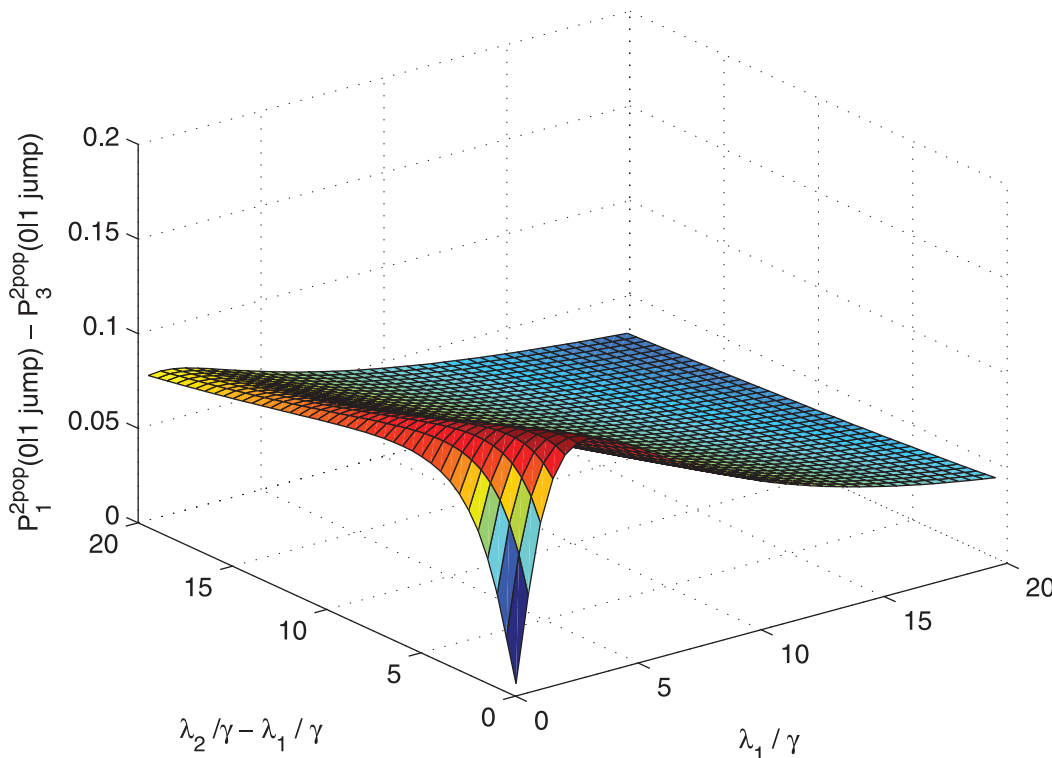


Figure 2. Comparison of probabilities of no secondary cases in a two-population model where individuals move only once. Variation of $P_1^{2pop}(0|1) - P_3^{2pop}(0|1)$ is represented as a function of λ_2/γ and λ_1/γ which vary on plausible ranges of values, under the constraint $\lambda_2/\gamma \geq \lambda_1/\gamma$. All parameters and variables are explained in Table 1. doi:10.1371/journal.pone.0009371.g002

For any z in $(0,1)$, $G_n(z) \leq G_{n+1}(z)$ (the inequality is strict unless $z = 1$). This is easily shown by considering the continuous function $G(x) = \left(\frac{1}{1+(R(1-z))/x}\right)^x$ which has a negative derivative $G'(x) = G(x) \left(\frac{R(1-z)}{R(1-z)+x} + \ln \frac{x}{R(1-z)+x}\right)$ for any $z \in (0,1)$. Moreover, $\lim_{n \rightarrow \infty} G_n(z) = e^{-R(1-z)}$ which corresponds to the p.g.f. of X_r when the infectious sojourn time is constant. As $P_n(0) = G_n(0)$, we have, $\forall n > 1$:

$$P_1(0) > P_n(0) > P_\infty(0). \tag{11}$$

This inequality implies that, for a given individual, the probability of producing no secondary infections is the greatest if the sojourn time in the infectious compartment follows an exponential distribution, which means that in this case the infection is more prone to extinction in early phases.

An order relation can be also established on $P_{n,ext}(g)$ defined in eq. (7). As $P_{n,ext}(0) = P_n(0)$, following eq. (11), we have more generally that $P_{n,ext}(0) \geq P_{n+1,ext}(0)$, $\forall n \geq 1$. By recurrence, we can show that this inequality holds for all g . We assume that this statement is true for $g-1$ and will prove that it is also true for g . Indeed, by using the fact that, for any z in $(0,1)$, $G_n(z)$ is strictly decreasing in n (unless $z = 1$) we have that: $P_{n,ext}(g) = G_n(P_{n,ext}(g-1)) \geq G_n(P_{n+1,ext}(g-1)) \geq$

$G_{n+1}(P_{n+1,ext}(g-1)) = P_{n+1,ext}(g)$. This inequality, suggesting that the exponential model drives more often to extinction even in long term, also holds for the ultimate extinction (whose probability is, for a Galton-Watson process, equal to the smallest positive root of $G_n(z) = z$). This corroborates the well known result postulating the decrease of the extinction probability as n increases.

Comparison of Probabilities of No Secondary Cases in a System of Two Coupled Populations with Respect to the Infectious Period Distribution

The main scope of this section is to derive a simpler expression for eq. (6), at least for special cases. In a more general context, Takacs [29] provided the distribution of the total time spent in one given state during the time interval $(0,u)$ for a process which is assumed alternating between two states. The case where the process is Markovian is treated as a particular case in [29]. Then, $g(u,\tau)$ can be calculated as $\partial_t \Omega(u,u-\tau)$, where $\Omega(u,\tau) = e^{-\sigma_1(u-\tau)} \left(1 + \sqrt{\sigma_1\sigma_2(u-\tau)} \int_0^\tau e^{-\sigma_2 y} y^{-1/2} I_1(2\sqrt{\sigma_1\sigma_2(u-\tau)y}) dy\right)$ is given in [29], with $I_1(x) = \sum_{j=0}^\infty \frac{(x/2)^{2j+1}}{j!(j+1)!}$ the modified Bessel function of the first kind of order 1.

Instead of differentiating $\Omega(u,u-\tau)$ we prefer to provide a more direct and intuitive way to calculate $g(u,\tau)$. If the number of jumps from one population to the other, which can be odd $(2k+1)$ or

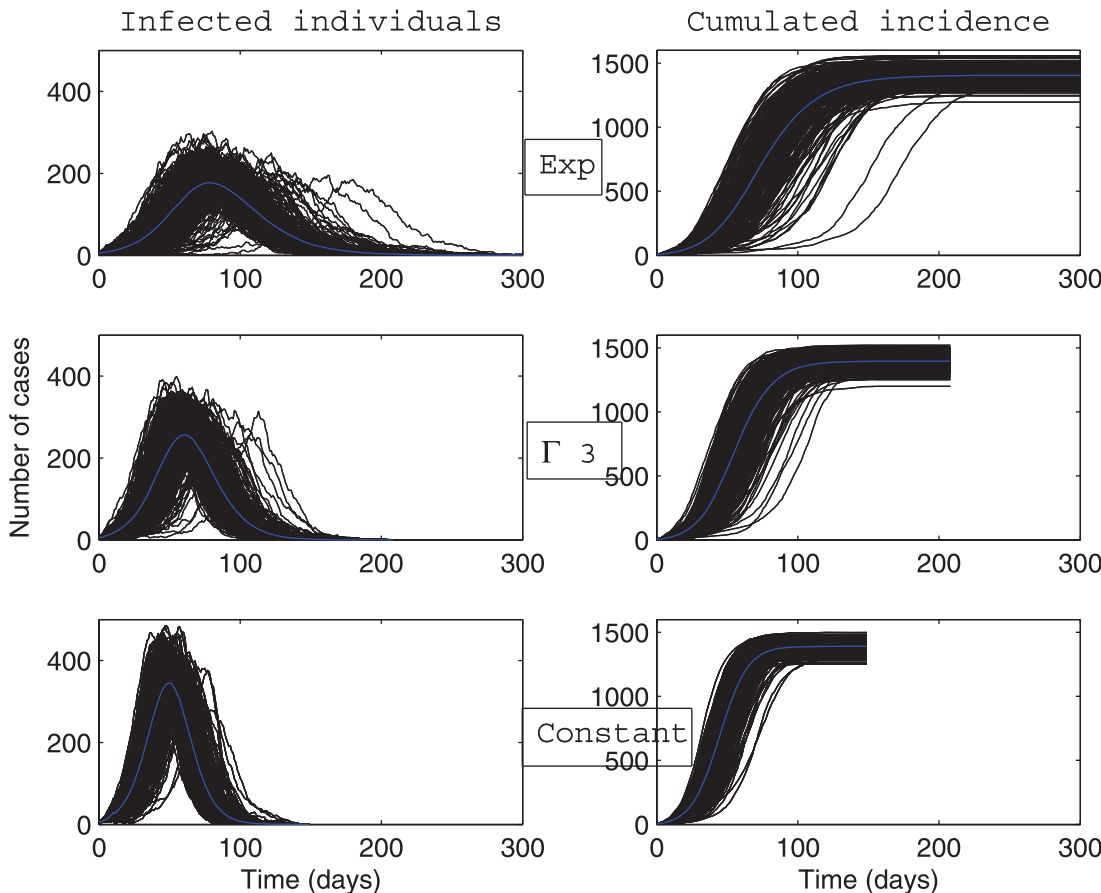


Figure 3. Dynamics of global number of cases and cumulative incidence described by a stochastic metapopulation model based on a completely connected network. Early extinct trajectories were not considered. The mean (blue curve) was calculated over major epidemics only (corresponding to a final attack rate greater than 5%). Simulations are performed using a time-continuous event-driven approach with $T_I \sim \text{Exp}(\gamma)$ (top panel), $T_I \sim \Gamma(3,3\gamma)$ (middle panel) and $T_I = \text{Const} = 1/\gamma$ (bottom panel). Parameters values are given in the subsection Examples of Results. doi:10.1371/journal.pone.0009371.g003

even $(2k)$, is taken into account, we can write $g(u, \tau) = \sum_{k=1}^{\infty} g_{2k}(u, \tau) + \sum_{k=0}^{\infty} g_{2k+1}(u, \tau)$. Let us explain the case where the number of switches between populations is odd ($=2k+1$), the individual spending $k+1$ time intervals in each patch. If the respective durations of these time intervals are $\tau_{1,j}$ and $\tau_{2,j}$ ($j=1, \dots, k+1$), their joint probability density function is:

$$\begin{aligned} f_{\tau}(\tau_{1,1}, \tau_{2,1}, \tau_{1,2}, \tau_{2,2}, \dots, \tau_{1,k+1}, \tau_{2,k+1}) &= \sigma_1 e^{-\sigma_1 \tau_{1,1}} \sigma_2 e^{-\sigma_2 \tau_{2,1}} \dots \sigma_1 e^{-\sigma_1 \tau_{1,k+1}} e^{-\sigma_2 \tau_{2,k+1}} \\ &= \sigma_1^{k+1} e^{-\sigma_1 \sum_{j=1}^{k+1} \tau_{1,j}} \sigma_2^k e^{-\sigma_2 \sum_{j=1}^{k+1} \tau_{2,j}} \\ &= \sigma_1^{k+1} e^{-\sigma_1 \tau} \sigma_2^k e^{-\sigma_2 (u-\tau)}. \end{aligned}$$

$g_{2k+1}(u, \tau)$ is obtained by integrating over all possible values of $\tau_{1,j}$ and $\tau_{2,j}$:

$$\begin{aligned} &\int_0^{\tau} \int_0^{u-\tau} \dots \int_0^{\tau-\tau_{1,1}-\dots-\tau_{1,k-1}} \int_0^{u-\tau-\tau_{2,1}-\dots-\tau_{2,k-1}} f_{\tau}(x_{1,1}, x_{1,2}, \dots, x_{1,k}, x_{2,k}) dx_{1,1} dx_{1,2} \dots dx_{1,k} dx_{2,k} \\ &= \sigma_1 \frac{(\sigma_1 \tau)^k}{k!} e^{-\sigma_1 \tau} \frac{(\sigma_2 (u-\tau))^k}{k!} e^{-\sigma_2 (u-\tau)}. \end{aligned} \tag{12}$$

The case where the number of jumps is even is inferred in a similar manner:

$$g_{2k}(u, \tau) = \frac{(\sigma_1 \tau)^k}{k!} e^{-\sigma_1 \tau} \sigma_2 \frac{(\sigma_2 (u-\tau))^{k-1}}{(k-1)!} e^{-\sigma_2 (u-\tau)}. \tag{13}$$

Finally, by summing up eqs. (12) and (13) over all values of k we obtain:

$$\begin{aligned} g(u, \tau) &= e^{-\sigma_1 \tau - \sigma_2 (u-\tau)} \left(\frac{\sqrt{\sigma_1 \sigma_2 \tau (u-\tau)}}{u-\tau} I_1(2\sqrt{\sigma_1 \sigma_2 \tau (u-\tau)}) + \sigma_1 I_0(2\sqrt{\sigma_1 \sigma_2 \tau (u-\tau)}) \right). \end{aligned} \tag{14}$$

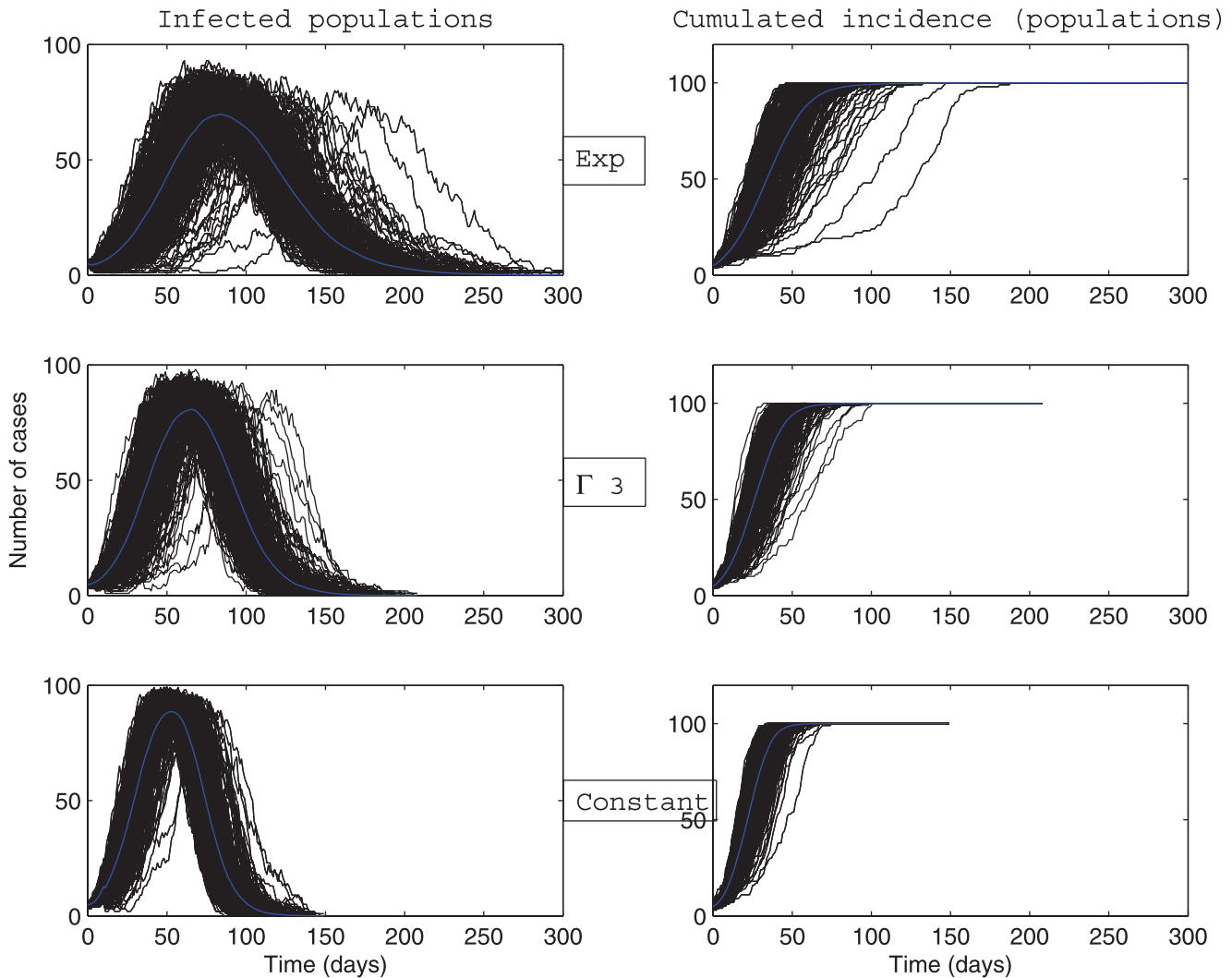


Figure 4. Dynamics of infected populations and cumulative incidence (in number of populations) described by a stochastic metapopulation model based on a completely connected network. Early extinct trajectories were not considered. The mean (blue curve) was calculated over major epidemics only (corresponding to a final attack rate greater than 5%). Simulations are performed using a time-continuous event-driven approach with $T_I \sim \text{Exp}(\gamma)$ (top panel), $T_I \sim \Gamma(3, 3\gamma)$ (middle panel) and $T_I = \text{Const} = 1/\gamma$ (bottom panel). Parameters values are given in the subsection Examples of Results. doi:10.1371/journal.pone.0009371.g004

where $I_0(x) = \sum_{j=0}^{\infty} \frac{(x/2)^{2j}}{j!j!}$ is the modified Bessel function of the first kind of order 0.

When replacing eq. (14) in eq. (6), the expression of $P_n^{2pop}(0)$ is too cumbersome for allowing the comparison with respect to different shapes of f_n . In order to circumvent this problem, the comparison is first explored numerically and then performed analytically on simplified but realistic cases.

We calculate $P_1^{2pop}(0)$ ($T_I \sim \text{Exp}(\gamma)$), $P_3^{2pop}(0)$ ($T_I \sim \Gamma(3, 3\gamma)$) and $P_{\infty}^{2pop}(0)$ ($T_I = 1/\gamma$) for 100 distinct sets of parameter values (for $\sigma_1, \sigma_2, \lambda_1$ and λ_2). Each value of $P_n^{2pop}(0)$ in Figure 1 represents the mean over 100000 Monte-Carlo simulations of a time continuous event-driven approach. According to our simulations, $P_n^{2pop}(0)$ decreases with n for most of the parameter combinations tested. Nonetheless, the graphic also suggests the existence of regions in the parameter space where $P_{\infty}^{2pop}(0) > P_1^{2pop}(0)$ (both probabilities being close to 1). The following two special cases illustrate each of these situations.

Let us consider the case where the individual cannot transmit the disease within his initial population ($\lambda_1 = 0$) where he cannot return ($\sigma_2 = 0$) once he left. If we take $\sigma_1 = \lambda_2 = \mu$, we have $P_1^{2pop}(0) = \frac{1 + 2\mu/\gamma}{(1 + \mu/\gamma)^2} < P_{\infty}^{2pop}(0) = e^{-\mu/\gamma}(1 + \mu/\gamma)$ for all $\mu/\gamma < r$, where r is the root of $P_1^{2pop}(0) - P_{\infty}^{2pop}(0) = 0$.

Another example practically relevant corresponds to a situation where the number of movements between patches is fixed and equal to one and there is the same probability to switch between patches ($\sigma_1 = \sigma_2 = \sigma$). This configuration could correspond to a metapopulation of farms, where animals do not change their original location, unless they are sold (or bought), events that generally occur only once during their lifetime. In this specific case, $g(u, \tau)$ is replaced by $g(u, \tau | 1 \text{ jump})$ and can be calculated using Bayes' formula as $g(u, \tau | 1 \text{ jump}) = \frac{g(u, \tau, 1 \text{ jump})}{\int_0^u g(u, \tau, 1 \text{ jump}) d\tau} = \frac{\sigma e^{-\sigma u}}{\int_0^u \sigma e^{-\sigma u} d\tau} = \frac{1}{u}$. When replacing this expression in eq. (6), where T_1 is reduced to its

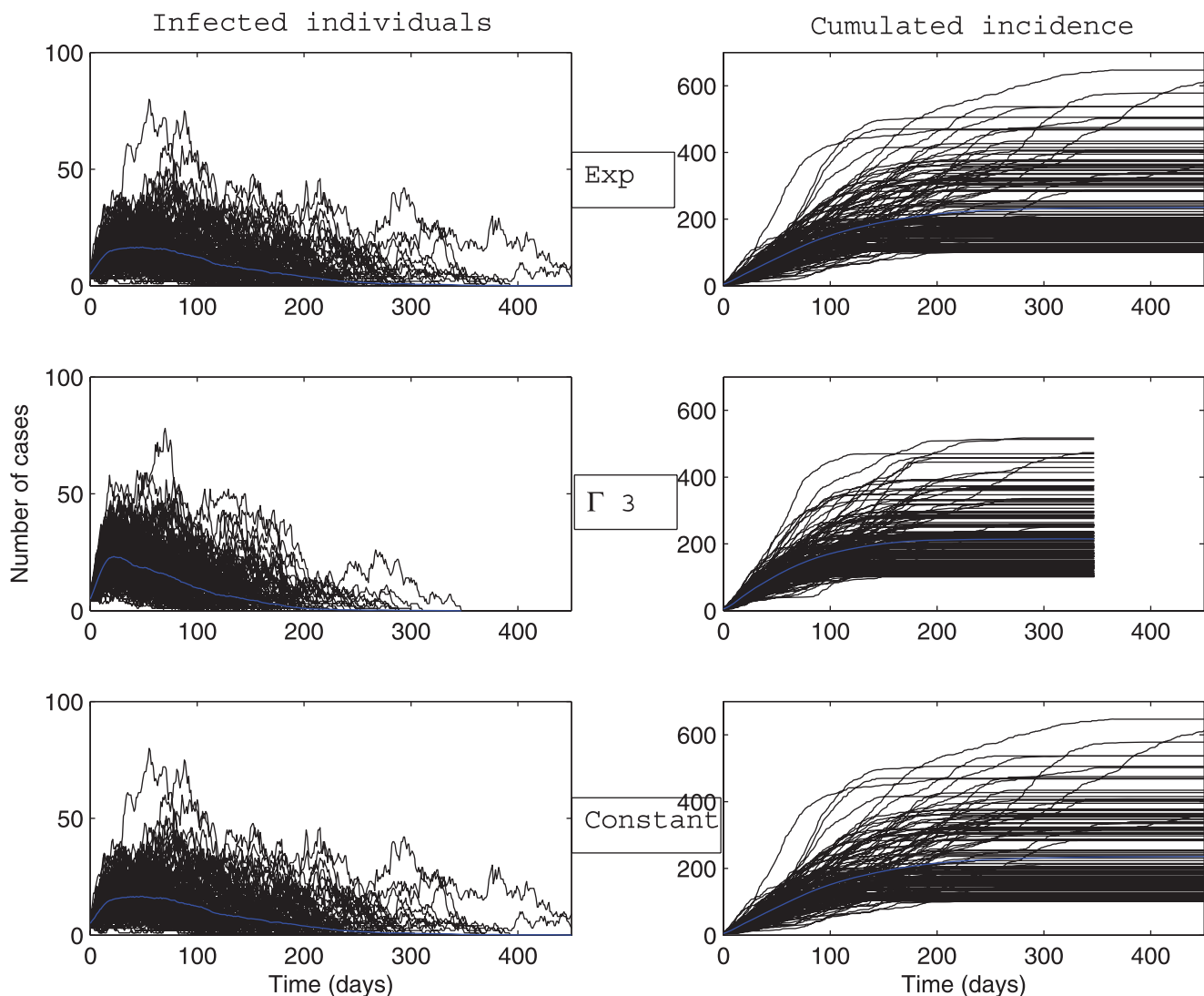


Figure 5. Dynamics of global number of cases and cumulative incidence described by a stochastic metapopulation model based on a scale-free network with mean degree of connectivity equal to 10. Early extinct trajectories were not considered. The mean (blue curve) was calculated over major epidemics only (corresponding to a final attack rate greater than 5%). Simulations are performed using a time-continuous event-driven approach with $T_I \sim \text{Exp}(\gamma)$ (top panel), $T_I \sim \Gamma(3, 3\gamma)$ (middle panel) and $T_I = \text{Const} = 1/\gamma$ (bottom panel). Parameters values are given in the subsection Examples of Results.

doi:10.1371/journal.pone.0009371.g005

continuous component only, we obtain:

$$P_{\infty}^{2pop}(0|1 \text{ jump}) = \frac{e^{-\lambda_1/\gamma} - e^{-\lambda_2/\gamma}}{(\lambda_2 - \lambda_1)/\gamma},$$

$$P_1^{2pop}(0|1 \text{ jump}) = \frac{\ln(\lambda_2/\gamma + 1) - \ln(\lambda_1/\gamma + 1)}{(\lambda_2 - \lambda_1)/\gamma}, \tag{15}$$

$$P_n^{2pop}(0|1 \text{ jump}) = \frac{n^n}{(n-1)(\lambda_2 - \lambda_1)/\gamma} \left(\frac{1}{(\lambda_1/\gamma + n)^{n-1}} - \frac{1}{(\lambda_2/\gamma + n)^{n-1}} \right).$$

$P_n^{2pop}(0|1 \text{ jump})$ can also be viewed as the mean of the random variable $h(T_I)$, where $h(u) = \frac{e^{-\lambda_1 u} - e^{-\lambda_2 u}}{u(\lambda_2 - \lambda_1)}$. Indeed,

$$P_n^{2pop}(0|1 \text{ jump}) = \int_0^{\infty} \frac{(\gamma n)^n u^{n-2} e^{-u\gamma}}{(n-1)!} \frac{e^{-\lambda_2 u} - e^{-\lambda_1 u}}{u} du \left(\int_0^u e^{-\tau(\lambda_1 - \lambda_2)} d\tau \right) = \int_0^{\infty} \frac{(\gamma n)^n u^{n-2} e^{-u\gamma}}{(n-1)!} \frac{e^{-\lambda_1 u} - e^{-\lambda_2 u}}{u(\lambda_2 - \lambda_1)} du = \mathbb{E}[h(T_I)].$$
 According to

Jensen's inequality applied to the convex function $h(u)$, we obtain that $P_n^{2pop}(0|1 \text{ jump}) = \mathbb{E}[h(T_I)] \geq h(\mathbb{E}[T_I]) = P_{\infty}^{2pop}(0|1 \text{ jump})$. For λ_2/γ and λ_1/γ varying on a plausible range of values we have that $P_1^{2pop}(0|1 \text{ jump}) - P_3^{2pop}(0|1 \text{ jump}) > 0$, as illustrated in Figure 2. In this case $P_1^{2pop}(0|1 \text{ jump}) \geq P_3^{2pop}(0|1 \text{ jump}) \geq P_{\infty}^{2pop}(0|1 \text{ jump})$.

Contrary to single homogeneously mixing populations, where the probability of generating no secondary cases always decreases as n increases, in the case of two coupled populations the monotony of $P_n^{2pop}(0)$ depends on parameter values: it is increasing when $P_n^{2pop}(0)$ is close to 1 and decreasing elsewhere.

Examples of Simulated Epidemic Dynamics in a Metapopulation

As an example, here we present for comparison graphical results for a complete graph and a scale-free network with mean degree of connectivity equal to 10. These results illustrate the impact of the mean connectivity and the distribution of the degree of

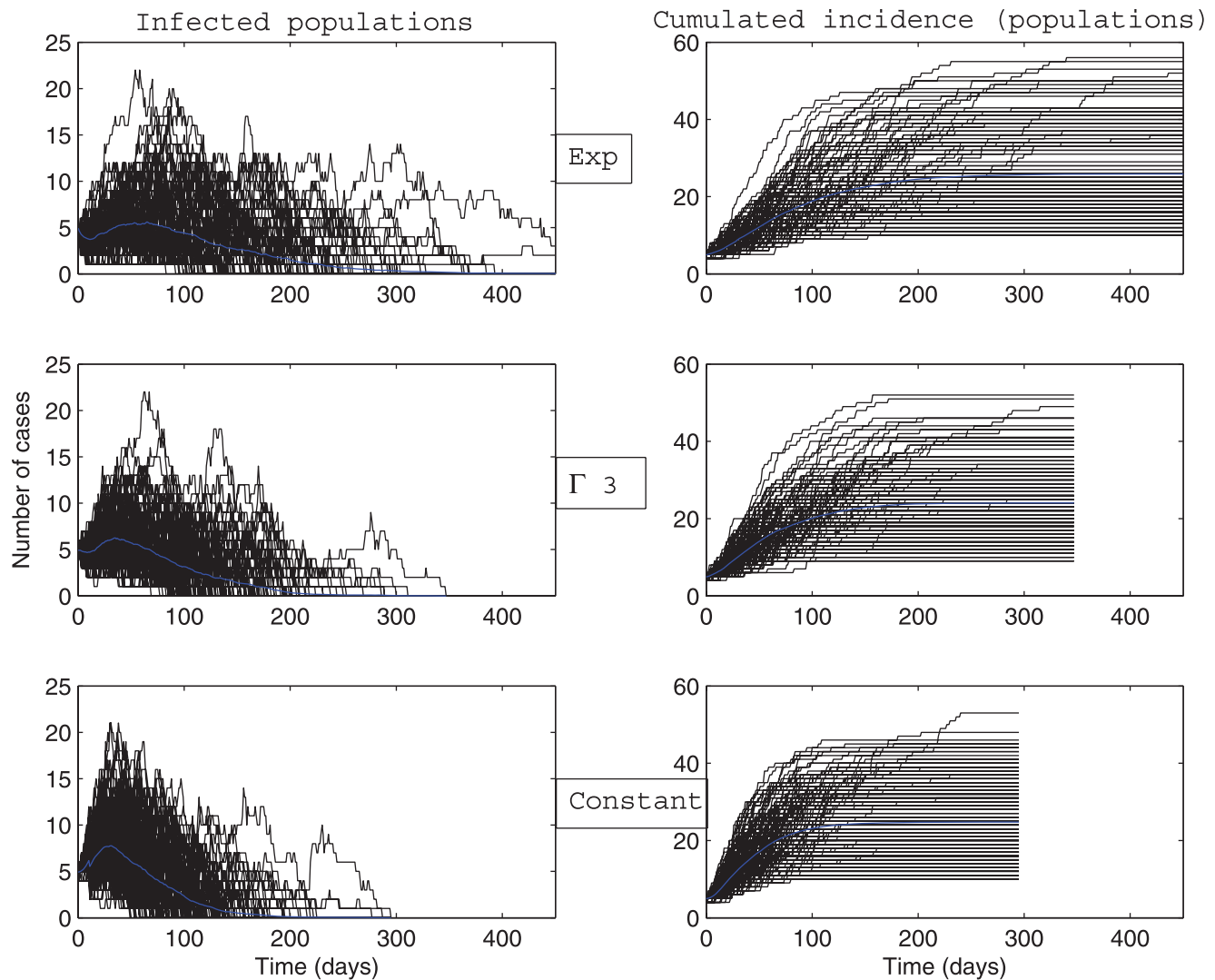


Figure 6. Dynamics of infected populations and cumulated incidence (in number of populations) described by a stochastic metapopulation model based on a scale-free network with mean degree of connectivity equal to 10. Early extinct trajectories were not considered. The mean (blue curve) was calculated over major epidemics only (corresponding to a final attack rate greater than 5%). Simulations are performed using a time-continuous event-driven approach with $T_I \sim \text{Exp}(\gamma)$ (top panel), $T_I \sim \Gamma(3, 3\gamma)$ (middle panel) and $T_I = \text{Const} = 1/\gamma$ (bottom panel). Parameters values are given in the subsection Examples of Results. doi:10.1371/journal.pone.0009371.g006

connectivity on the outputs of an epidemic spreading on the network, since the two networks selected are different with respect to these two factors. For each of the three models based on different infectious period distributions, $T_I \sim \text{Exp}(\gamma)$ (exponential), $T_I = 1/\gamma$ (constant) and $T_I \sim \Gamma(3,3\gamma)$ (gamma), with equal mean $1/\gamma = 10$ days (corresponding to $R_0 = 2$), simulations were performed for $p = 100$ populations of $N_i = 20$ individuals each, coupled by identical migration rates ($\sigma_{ij} = \sigma_{ji} = 0.001, \forall i, j$) and characterized by equal transmission rates ($\beta_i = 0.2 \text{day}^{-1}, \forall i$). Five of the 100 populations were randomly seeded with $I(0) = 1$.

Whatever the scenario considered (Figures 3, 4, 5, 6) and for the parameter values used here, there is a high variability in incidence at the individual and the population levels. Epidemic spread is slowed down by the random nature of the migration between populations. This effect is increased when the mean connectivity decreases (Figures 5, 6). Regardless of the network structure, noteworthy differences between dynamics with respect to the distribution of infectious sojourn time are obtained for the

amplitude and the date of the epidemic peak and the epidemic duration. Peaks of the mean incidence at individual and population levels decrease from $T_I = \text{Const} = 1/\gamma$ to $T_I \sim \Gamma(3,3\gamma)$ and to $T_I \sim \text{Exp}(\gamma)$. In contrast, the epidemic duration and the variability in incidence increase from the constant to the exponential distributed sojourn times in infectious status (left graphs of Figures 3, 4, 5, 6). For both network topologies in our example, the maximal epidemic duration roughly doubles between the constant and the exponential distributed infectious time based models. The main difference between the networks tested, which could potentially interfere with the influence of the infection time distribution, is illustrated in Figure 7. This figure represents the empirical distribution of the final epidemic size calculated on the 300 stochastic simulations for each scenario. The proportion of minor epidemics is significantly lower in the completely connected graph compared to the low connected scale-free network (right and left panels of Figure 7 respectively), since this latter network facilitates the occurrence of

Distribution of the final epidemic size

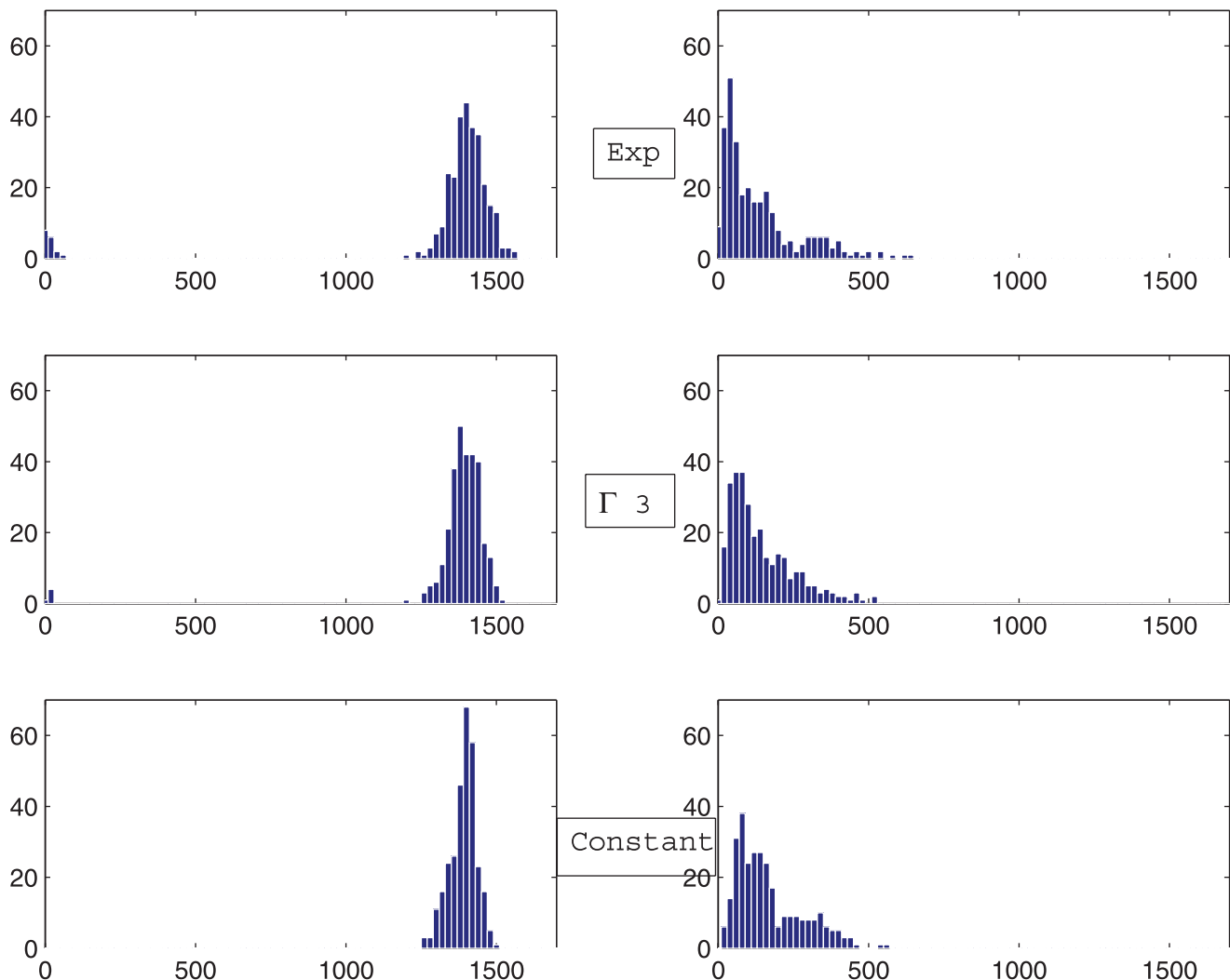


Figure 7. Distribution of the final epidemic size. Calculation was performed on 300 simulations of a stochastic metapopulation model based on a completely connected graph (left panel) and on a scale-free network with mean degree of connectivity equal to 10 (right panel), with $T_I \sim \text{Exp}(\gamma)$ (top graphs), $T_I \sim \Gamma(3,3\gamma)$ (middle graphs) and $T_I = \text{Const} = 1/\gamma$ (bottom graphs). Parameters values are given at page 10. doi:10.1371/journal.pone.0009371.g007

early fade-outs. Moreover, this probability decreases from exponentially distributed to constant infectious durations for both networks. This tendency is more marked for the scale-free network.

Sensitivity Analysis of the Outputs of Epidemic Dynamics in Metapopulations

As illustrated by simulations described in the previous section, the choice of the distribution of the infection duration has various impacts on global epidemic outputs such as epidemic duration, epidemic size or epidemic peak in a metapopulation. In addition, these impacts could vary with respect to other parameters. The sensitivity analysis provides a quantification of these potential interactions. Namely, we explore whether predominant contributing factors are the same for every output of the epidemic and also if they are identical for the three variants 'E-G', 'G-C' and 'E-C' of a given output. Results are summarized in Figure 8, 9 and 10: each bar is decomposed in segments with heights equal to the sensitivity coefficients, which evaluate the percentage of the global variability of the dependent variable explained by each factor.

In order to explicitly illustrate our sensitivity analysis approach, let us describe, as an example, the statistical analysis performed on the epidemic size. For a given set of parameters we simulate

epidemics with exponentially, gamma distributed and constant infectious periods and calculate means and variances of their epidemic sizes, noted by ME_{Sexp} , ME_{Sgamma} , ME_{Sconst} , VE_{Sexp} , VE_{Sgamma} , VE_{Sconst} respectively. To account for nonlinear relationships between outputs and factors we apply the log transformation to outputs. Then, a classical ANOVA is performed on each couple of differences, as dependent variables, $(\log(ME_{Sexp})-\log(ME_{Sgamma}))$, $(\log(ME_{Sgamma})-\log(ME_{Sconst}))$, $(\log(ME_{Sexp})-\log(ME_{Sconst}))$, $(\log(VE_{Sexp})-\log(VE_{Sgamma}))$, $(\log(VE_{Sgamma})-\log(VE_{Sconst}))$ and $(\log(VE_{Sexp})-\log(VE_{Sconst}))$. For each of these dependent variables, sensitivity coefficients are then calculated using the equation (8) for each of the factors of interest (Figures 8, 9, 10). Let us focus on differences between exponential and gamma based models. For instance, according to Figure 8, we can say that R_0 is the most influential factor impacting on the difference in the mean epidemic expressed in terms of infected individuals. In other words, the error in the prediction of the mean epidemic size committed if an exponential distribution was used instead of more realistic gamma distribution for the infectious duration heavily depends on R_0 . At a population scale, the most important factor for the epidemic size is the migration intensity, whereas differences in variances are almost equally and weakly impacted by all factors.

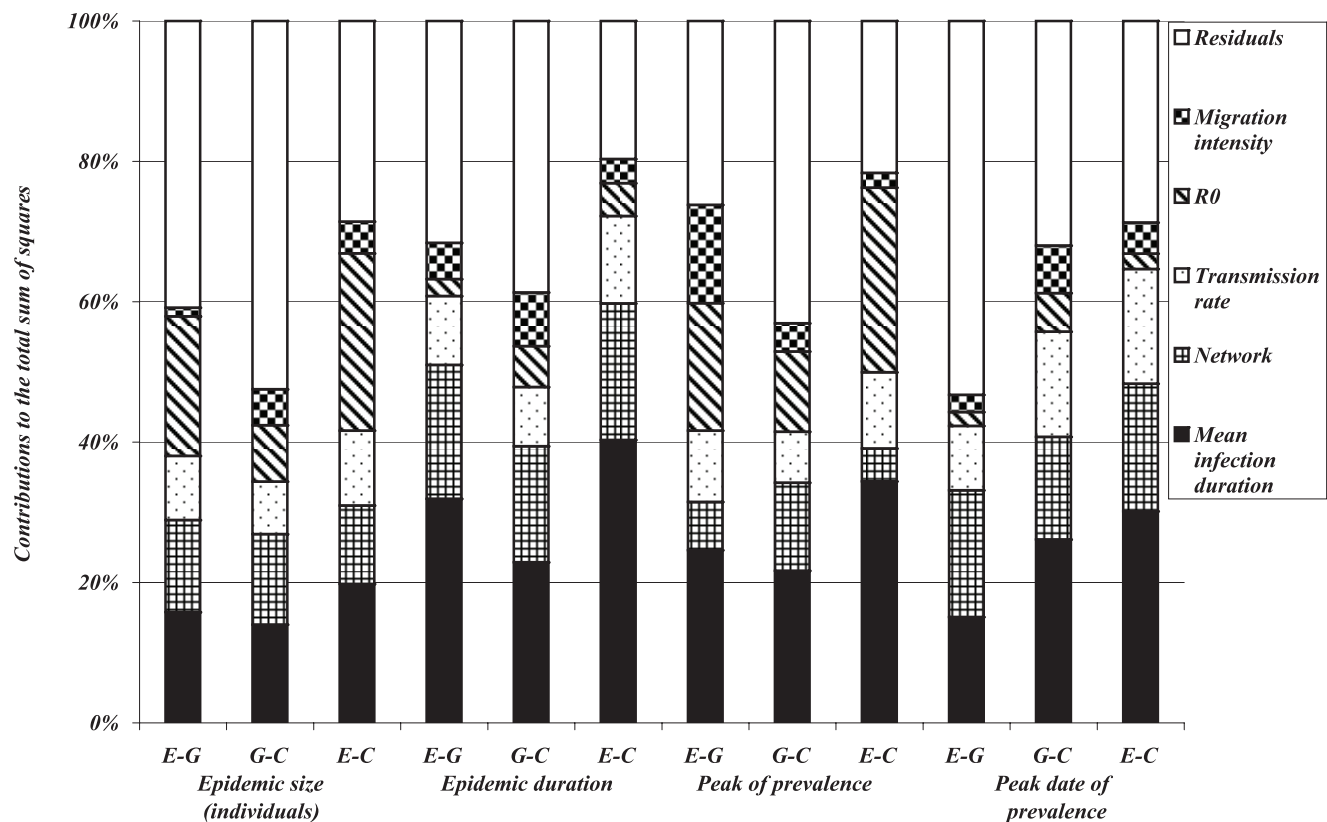


Figure 8. Results of ANOVA on 960 simulated scenarios of epidemic spread with parameter values given in Table 2. Dependent variables (on x-axis) are logarithm of means (over the non early extinct dynamics) of global variables (directly referring to individuals regardless of their population of origin): size and duration of the epidemic, size and date of the epidemic peak. For each of these outputs three variants are considered with respect to the distribution of infection duration: 'E-G' (difference between the value of the output simulated with the exponentially distributed infectious period and the value corresponding to the gamma distributed infectious period), 'G-C' (difference between the value of the output simulated with the gamma distributed infectious period and the value corresponding to a constant infectious period) and 'E-C' (difference between the value of the output simulated with the exponentially distributed infectious period and the value corresponding to a constant infectious period). Different pattern fills correspond to contributions of five input factors (mean infection duration, network, transmission rate, R_0 and migration intensity) to the variation in outputs amongst scenarios. doi:10.1371/journal.pone.0009371.g008

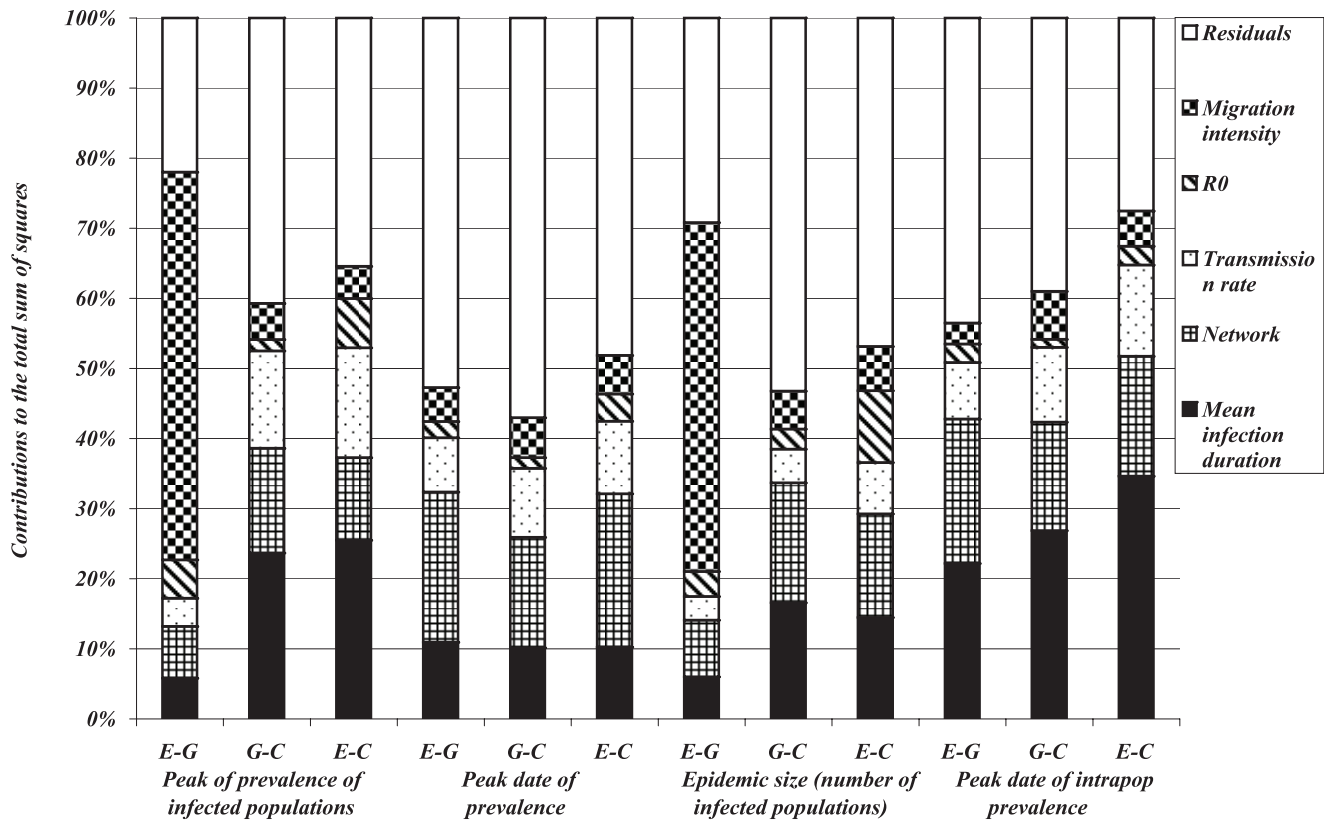


Figure 9. Results of ANOVA on 960 simulated scenarios of epidemic spread with parameter values given in Table 2. Dependent variables (on x-axis) are logarithm of means (over the non early extinct dynamics) of global variables (referring to populations): size and duration of the epidemic, size and date of the epidemic peak. For each of these outputs three variants are considered with respect to the distribution of infection duration: 'E-G' (difference between the value of the output simulated with the exponentially distributed infectious period and the value corresponding to the gamma distributed infectious period), 'G-C' (difference between the value of the output simulated with the gamma distributed infectious period and the value corresponding to a constant infectious period) and 'E-C' (difference between the value of the output simulated with the exponentially distributed infectious period and the value corresponding to a constant infectious period). Different pattern fills correspond to contributions of five input factors (mean infection duration, network, transmission rate, R_0 and migration intensity) to the variation in outputs amongst scenarios.

doi:10.1371/journal.pone.0009371.g009

More generally, when analyzing means over major epidemics of outputs directly referring to individuals, the main factors explaining their variations are first the mean infection duration and second the network structure or R_0 , depending on the output of interest (Figure 8). The migration intensity contributes the least to the majority of outputs. This order of importance is preserved for most outputs whatever the distributions compared ('E-G', 'G-C' and 'E-C'). For outputs directly referring to the population level (Figure 9), the mean infection duration and the network structure are the most important factors for the majority of outputs. As an exception, the migration intensity plays the major role for the 'E-G' variant of the epidemic size and of the peak of prevalence. The outputs expressed as variances over all simulated major epidemics (Figure 10) are less influenced by input factors. For the majority of outputs the most contributing factors are R_0 , the network structure, and the mean infection duration. It is noticeable that the variance of 'E-C' variant of nearly all outputs is better explained (higher coefficients of determination R^2) by input factors than 'G-C' and 'E-G' variants.

Regarding the early extinction, the most important correlations of the proportion of minor epidemics are with R_0 (PRCC = -0.57), the migration intensity (PRCC = -0.48) and the network (PRCC = -0.36). The metapopulation epidemic has more chances to go extinct as R_0 decreases and as populations are less

connected. The distribution of the infectious period is more weakly correlated with the proportion of early extinct dynamics (PRCC = 0.13).

Discussion

In this paper, we studied the impact of the infectious period distribution on the global dynamics described by a metapopulation model comprising many patches. Since assuming exponential sojourn times in infected states is a common approximation used in most of mathematical models, we were interested in evaluating the potential bias that such an assumption would introduce at a global scale of an epidemic. Besides some analytical developments extending the work of Keeling and Grenfell [3] in single populations, we deliberately focused our attention on metapopulations. There are two main reasons for this choice. First, the impact of the infection period distribution in single populations was extensively explored in the literature ([2,3] and many other studies, especially in the context of HIV epidemic). Second, we were interested in exploring the potential interactions between effects of the infection duration distribution and of exogenous factors (such as the topology of the network underlying the metapopulation, or the migration intensity), elements which are not present in single populations.

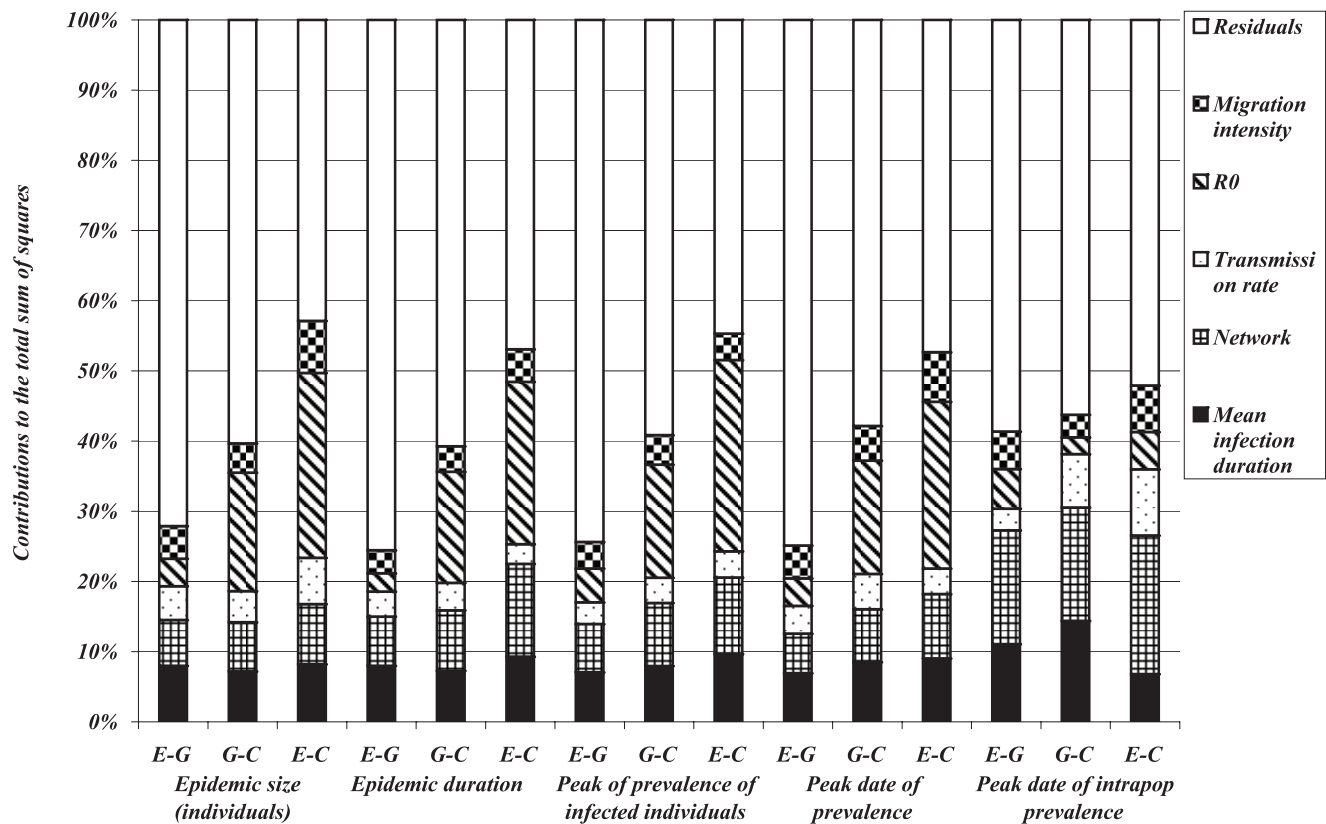


Figure 10. Results of ANOVA on 960 simulated scenarios of epidemic spread with parameter values given in Table 2. Dependent variables (on x-axis) are logarithm of variances (over the non early extinct dynamics) of global variables (directly referring to individuals regardless of their population of origin): size and duration of the epidemic, size and date of the epidemic peak and date of intra-population epidemic peak. For each of these outputs three variants are considered with respect to the distribution of infection duration: 'E – G' (difference between the value of the output simulated with the exponentially distributed infectious period and the value corresponding to the gamma distributed infectious period), 'G – C' (difference between the value of the output simulated with the gamma distributed infectious period and the value corresponding to a constant infectious period) and 'E – C' (difference between the value of the output simulated with the exponentially distributed infectious period and the value corresponding to a constant infectious period). Different pattern fills correspond to contributions of five input factors (mean infection duration, network, transmission rate, R_0 and migration intensity) to the variation in outputs amongst scenarios. doi:10.1371/journal.pone.0009371.g010

First, we performed analytical investigations. As mentioned above, we extended developments of [3] for single populations to the case of gamma distributed infectious durations, and showed that the probability of extinction in early phases increases with the variance of the infectious duration (which is the largest for exponential distribution, given the same mean for all distributions and $n \geq 1$ for $\Gamma(n, m)$). This result corroborates what it was shown for branching-type epidemic processes with an $R_0 > 1$ [2]. The same order in the extinction probabilities holds after g generations of infecteds.

The case of two coupled populations was also explored through a similar approach invoking Poisson processes for the generation of secondary infection and Markovian transitions for the migration between patches. Under the constraint that an individual changes only one time his patch of residence during the duration of his infectious period, a plausible assumption for real animal populations, we showed that here again the probability of producing no secondary cases is the greatest when T_I is exponentially distributed. Inversely, we identified special cases where the probability of early extinction in a two-population system increases as n increases. This occurs, for instance, under a specific constraint on parameters, in a situation where the individual cannot transmit the disease within his initial population where he cannot return once he left.

Therefore, caution has to be paid when interpreting extinction results provided by mathematical models: the extinction probability could be over or underestimated in situations where the exponential distribution is particularly not appropriate for modelling the infection duration. This could be the case for long lasting diseases where the probability of recovering strongly depends on the disease stage.

Although analytical explorations were performed on *SIR* models, they would be completely transposable to the *SEIR* case, as the introduction of a latent period would not affect the quantities analyzed in this study. Besides, some of our results, as those in [3], were rigorously proved based on the assumption of no depletion of susceptibles during the epidemic process. Since this is an acceptable approximation when sizes of target populations are large, it could become awkward to defend it when working with small populations.

In a second part, we focused on a metapopulation context comprising more than two patches and relatively small population sizes, without making the assumption of constant proportion of susceptibles over time. Since analytical explorations were too complex for this case, it was studied by simulations and statistical analysis of simulated data. As already mentioned above, in the metapopulation framework, we were not exclusively interested in

the potential effect of the shape of the infection duration distribution on epidemic dynamics. We also explored in which extent this effect could be influenced by other factors, especially those directly related to the metapopulation structure, such as the mean degree of connectivity of the network or the migration intensity between patches.

The impact of network topology on the spread of epidemics was previously explored (see [30] for a review). However, the assumption for the distribution of infection period was neither discussed nor taken into account in those studies. In addition, most often, the nodes of such networks are individuals and not populations and infection spread between patches is indirectly considered through distance-based transmission rates. In our study, both intra-population epidemic dynamics (described by a specific *SIR* model for each patch) and inter-population spread of infection (taken into account through migration of individuals) were considered.

The sensitivity analysis performed indicated that the most important factors which influence the impact of infection duration distribution on epidemic outputs in metapopulations are R_0 , the network structure and the mean infection period. This means that under or overestimation of epidemic outputs such as epidemic duration and prevalence peak size, due to specific modelling choices concerning the distribution of infection duration, depends on other factors related to the infectious potential of the pathogen and to the way the populations are connected. Nevertheless, these effects are not identical for all criteria: for instance, the peak of population prevalence is strongly influenced by the migration intensity (Figure 9), whereas variation in its date is mostly explained by the network topology (Figure 10). A similar statistical analysis conducted within a single population and including only the transmission rate and the mean infections duration as factors (results not shown) reveals that R_0 is the most influential factor on size-related outputs (such as epidemic burden or epidemic peak), whereas the mean infection duration preferentially impacts time-related outputs (such as epidemic duration or peak date).

The sensitivity indices and hence the relative importance of each factor in explaining the variability of a given criterion could

depend on the number of factors included in the analysis, but their relative ordering should not change.

As a first analysis of the potential effect of the shape of infection duration distribution on global dynamics, all intra-population parameters have been considered consistent among populations of the network. We also considered small intra-patch population sizes (equal to 20). This could correspond to a group of small farms but also to the classes of a school and hence refers to human populations. Further research is needed to investigate heterogeneous metapopulations characterized by unequal contributions of patches to the global spread of a disease and also to rigorously assess the robustness of results to changes in the population size.

We have shown that the effect of assuming exponential distribution for infection periods instead of more realistic distributions varies with respect to the output of interest and to other exogenous factors. Attention has to be paid to all these elements in practice. For example, when evaluating control strategies at a global level by modelling approach in order to optimize their use, expected losses due to the disease spread as well as expected gains due to interventions may be misestimated. Future research should more closely explore under what circumstances the exact distributions are relevant in order to assess in which situation effort should be put into obtaining them.

Acknowledgments

The authors thank Jean-Baptiste Denis and Hervé Monod for their helpful comments on the sensitivity analysis and Pierre Nicolas, Valeria Banica and Thomas Duyckaerts for their valuable suggestions on the analytical developments.

Author Contributions

Conceived and designed the experiments: EV PE. Performed the experiments: EV HB. Analyzed the data: EV HB PE. Wrote the paper: EV PE.

References

- Keeling MJ, Rohani P (2008) Modeling infectious diseases in humans and animals. Princeton: Princeton University Press, first edition. 408 p.
- Malice MP, Kryscio RJ (1989) On the role of variable incubation periods in simple epidemic models. *Math Med Biol* 6: 233–242.
- Keeling MJ, Grenfell BT (1998) Effect of variability in infectious period on the persistence and spatial spread of infectious diseases. *Math Biosci* 147: 207–226.
- Wearing HJ, Rohani P, Keeling JM (2005) Appropriate models for the management of infectious diseases. *PLoS Med* 2: e174.
- Feng Z, Xu D, Zhao H (2007) Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bull Math Biol* 69: 1511–1536.
- Keeling MJ, Grenfell BT (1997) Disease extinction and community size: modeling the persistence of measles. *Science* 275: 65–67.
- Lloyd AL (2001) Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* 60: 59–71.
- Lloyd AL (2001) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc Biol Sci* 268: 985–993.
- Keeling MJ, Grenfell BT (2002) Understanding the persistence of measles: reconciling theory, simulation and observation. *Proc Biol Sci* 269: 335–343.
- Roberts MG, Heesterbeek JAP (2007) Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J Math Biol* 55: 803–816.
- Chowell G, Nishiura H (2008) Quantifying the transmission potential of pandemic influenza. *Physics of Life Reviews* 5: 50–77.
- Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 274: 599–604.
- Yan P (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *J Theor Biol* 251: 238–252.
- Sattenspiel L, Dietz K (1995) A structured epidemic model incorporating geographic mobility among regions. *Math Biosci* 128: 71–91.
- Fulford GR, Roberts MG, Heesterbeek JAP (2002) The metapopulation dynamics of an infectious disease: tuberculosis in possums. *Theor Popul Biol* 61: 15–29.
- Arino J, van den Driessche P (2003) A multi-city epidemic model. *Math Pop Stud* 10: 175–193.
- Lloyd AL, Jensen VAA (2004) Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. *Math Biosci* 188: 1–16.
- Hsieh YH, van den Driessche P, Wang L (2007) Impact of travel between patches for spatial spread of disease. *Bull Math Biol* 69: 1355–1375.
- McCormack RK, Allen LJS (2007) Multi-patch deterministic and stochastic models for wildlife diseases. *J Biol Dyn* 1: 63–85.
- Ferguson NM, Galvani AP, Bush RM (2003) Ecological and immunological determinants of influenza evolution. *Nature* 422: 428–433.
- Flahault A, Vergu E, Coudeville L, Grais RF (2006) Strategies for containing a global influenza pandemic. *Vaccine* 24: 6751–6755.
- Koelle K, Cobey S, Grenfell BT, Pascual M (2006) Epochal evolution shapes the phylodynamics of inter-pandemic influenza A (H3N2) in humans. *Science* 314: 1898–1903.
- Viboud C, Bjornstad ON, Smith DL, Simonsen L, Miller MA, et al. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 312: 447–451.
- Jesse M, Ezanno P, Davis S, Heesterbeek JAP (2008) A fully coupled, mechanistic model for infectious disease dynamics in metapopulations: movements and epidemic duration. *J Theor Biol*.
- Balenghien T, Chalvet-Montfray K, Lesnoff M, Thiaucourt F, Sabatier P, et al. (2004) Time-delay dynamics for contagious bovine pleuropneumonia. *Acta Biotheor* 52: 365–377.
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45: 167–256.
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, et al. (2008) Global sensitivity analysis, first edition John Wiley. 292 p.

28. Alvarez FP, Crépey P, Barthélemy M, Valleron A (2007) sispread: A software to simulate infectious diseases spreading on contact networks. *Methods Inf Med* 46: 19–26.
29. Takacs L (1957) On certain sojourn time problems in the theory of stochastic processes. *Acta Math Acad Sci Hungary* 8: 169–191.
30. Keeling MJ, Eames KTD (2005) Networks and epidemic models. *J R Soc Interface* 2: 295–307.