

## PUBLISHED VERSION

Ross, Joshua Vincent; House, Thomas; Keeling, Matt J..  
Calculation of disease dynamics in a population of households, *PLoS One*, 2010; 5(3):1-9.

Copyright 2010 Ross 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.

### PERMISSIONS

<http://www.plos.org/journals/license.php>

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.

The Public Library of Science (PLoS) applies the Creative Commons Attribution License (CCAL) to all works we publish (read the human-readable summary or the full license legal code). Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLoS journals, so long as the original authors and source are cited. No permission is required from the authors or the publishers.

2<sup>nd</sup> May 2011

<http://hdl.handle.net/2440/66428>

# Calculation of Disease Dynamics in a Population of Households

Joshua V. Ross<sup>1\*</sup>, Thomas House<sup>2</sup>, Matt J. Keeling<sup>2</sup>

**1** King's College, University of Cambridge, Cambridge, United Kingdom, **2** Biological Sciences and Mathematics Institute, University of Warwick, Coventry, United Kingdom

## Abstract

Early mathematical representations of infectious disease dynamics assumed a single, large, homogeneously mixing population. Over the past decade there has been growing interest in models consisting of multiple smaller subpopulations (households, workplaces, schools, communities), with the natural assumption of strong homogeneous mixing within each subpopulation, and weaker transmission between subpopulations. Here we consider a model of *SIRS* (susceptible-infectious-recovered-susceptible) infection dynamics in a very large (assumed infinite) population of households, with the simplifying assumption that each household is of the same size (although all methods may be extended to a population with a heterogeneous distribution of household sizes). For this households model we present efficient methods for studying several quantities of epidemiological interest: (i) the threshold for invasion; (ii) the early growth rate; (iii) the household offspring distribution; (iv) the endemic prevalence of infection; and (v) the transient dynamics of the process. We utilize these methods to explore a wide region of parameter space appropriate for human infectious diseases. We then extend these results to consider the effects of more realistic gamma-distributed infectious periods. We discuss how all these results differ from standard homogeneous-mixing models and assess the implications for the invasion, transmission and persistence of infection. The computational efficiency of the methodology presented here will hopefully aid in the parameterisation of structured models and in the evaluation of appropriate responses for future disease outbreaks.

**Citation:** Ross JV, House T, Keeling MJ (2010) Calculation of Disease Dynamics in a Population of Households. *PLoS ONE* 5(3): e9666. doi:10.1371/journal.pone.0009666

**Editor:** Mike B. Gravenor, University of Swansea, United Kingdom

**Received:** November 16, 2009; **Accepted:** February 17, 2010; **Published:** March 18, 2010

**Copyright:** © 2010 Ross 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 has been supported by funding from the Medical Research Council (Grant Number G0701256; www.mrc.ac.uk), and the Wellcome Trust (www.wellcome.ac.uk). 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: jvr25@cam.ac.uk

## Introduction

The earliest models proposed for infectious disease dynamics assumed that the population afflicted by the pathogen was large and homogeneously mixed such that deterministic equations with simple frequency-dependent transmission were appropriate [1–3]. These models were subsequently extended in three major directions: taking into account the discrete nature of populations and the stochastic nature of transmission and recovery [4–6]; taking into account heterogeneity between individuals in terms of differential mixing [7]; and accounting for spatial structure and the often localized transmission of infection (a variety of approaches are summarised in [3, Chapter 7]). This latter extension has taken several different forms but can be predominately dichotomized into those based upon explicit contact networks that determine the possible opportunities for transmission between individuals [8–17] and those models that stratify the population into sub-populations (for example households), with, typically, homogeneous mixing within the subpopulations and weaker mixing between them [18–25] (although there exist exceptions, including those that bridge this divide; see for example [26–28]). The former network models are most appropriate for situations where there exists explicit knowledge of contact structure heterogeneity such as for sexually transmitted diseases [13], the air transport network for SARS [29] and for livestock movements in Britain [30,31]. The latter structured or metapopulation models reflect the relatively strong opportunity for transmission between individuals within a

household compared to transmission to other individuals in the population. These households models can therefore be conceptualised as a combination of a network model (for the strong within household transmission) and a homogeneous mixing model (for the weaker transmission to the general population).

Household structured models offer an attractive trade-off between fine-scale detail and computational feasibility, and for this reason they increasingly form the basis of studies in disease management. For example, it has been known for some time that household structure influences the critical threshold for invasion and vaccination [20,32], and more recently ideas from formal modelling have been used to answer increasingly applied questions: these range from parameter estimation [33] to evaluation of an appropriate response to an influenza pandemic [34,35].

Here we consider a pathogen for which individuals develop transient immunity following infection, with the immunity waning resulting in the individual eventually returning to full-susceptibility to the disease: the *SIRS* (susceptible-infected-recovered-susceptible) model. We model the spread of such a pathogen amongst individuals occupying a very large (assumed infinite) set of households. The infection dynamics within each household is captured by a Markov chain representation, accounting for the stochastic effects due to the small number of individuals within each household. We assume that transmission between households results from (effective) contacts between individuals in the population occurring at a fixed rate. Models of this form have been studied widely recently [21–24,36,37], although the particular characteristic

of waning immunity has received relatively less attention [24,38]. We then extend this model to additionally incorporate more realistic distributions for the infectious period; this is achieved by using the classical approach of breaking the infectious class into a number of new classes while preserving the total average time in these new classes (see, for example, [39] and [40]).

In this paper we investigate a range of epidemiologically important quantities over a large region of parameter space; as such we present and utilise efficient methods for evaluating these quantities. Some of the methods we adopt have been described elsewhere, and alternative methods exist for evaluating several of the quantities. In particular the classic work on households models such as [21] often includes more general features than we consider here; however as a consequence of this generality, many of the results are not as readily amenable to rapid numerical evaluation. Here we present results and numerical methods appropriate for studying our model, and closely related ones, in a unified and accessible form, hence allowing their direct application to epidemiological problems. For this very reason we provide MATLAB code to evaluate the quantities considered (see Supporting Information File S2); MATLAB provides an ideal language and framework for the manipulation of the transition matrices that are associated with the Markov chain approach adopted here.

Five epidemiologically important quantities are evaluated: (i) the threshold for invasion,  $R_*$ , which is the household basic reproduction number [21], which measures the average number of secondary households infected from a single infected household in a totally susceptible population and therefore determines if a pathogen may successfully invade ( $R_* > 1$ ); (ii) the early growth rate,  $r$ , (also called the *Malthusian parameter* [21]), which is the quantity generally measured for statistical inference in the early stages of an emerging infection; (iii) the household offspring distribution, defined as the distribution of secondary households infected from an epidemic seeded with a single infectious individual – the expectation of this random variable is  $R_*$ ; (iv) the endemic prevalence of infection  $J^*$  amongst the community of households; and (v) the distribution of transient dynamics of the process. This latter calculation not only provides an opportunity to investigate the impact of the models structure in detail, but provides an alternative method to assess the accuracy of the many quantities listed above.

In the next section we introduce the basic model of within- and between-household transmission, before detailing the methodology adopted, and illustrate several of the techniques with respect to a model for households of size one (equivalent to the classical homogeneously-mixing *SIR* model). We then proceed to apply this methodology to our structured households models to investigate the impact of household size, within- and between-household transmission rates, rate of waning immunity, and infectious period distribution on dynamics. We conclude by discussing the general implications of our results.

## Methods

The household dynamics are described by three basic processes: transmission of infection between an infectious and susceptible individual within the household, recovery of infected individuals, and loss of immunity (Table 1). Recovery of an infected individual and loss of immunity for a recovered individual are both assumed to occur independently of the states of other individuals within the household with constant probabilistic rates  $\gamma$  and  $\mu$ , respectively. Transmission within the household is assumed to be frequency dependent with transmission parameter  $\beta$ ; note that the  $(N-1)$  term in the denominator is to ensure frequency dependent contact with all other members of the household. In situations where we need to consider the transient or long-term dynamics it becomes important to allow infection to enter the household from the external population. This is captured by an external force of infection  $\varepsilon$  to all susceptible individuals within the household. Finally, we assume a between-household transmission rate  $\alpha$ , such that each infectious individual within the household generates secondary cases at rate  $\alpha$ . Clearly, for self-consistent dynamics we insist that  $\varepsilon = \alpha J(t)$ , where  $J(t)$  is the proportion of infectious individuals in the population.

The first four quantities of interest ( $R_*$ ,  $r$ , distribution of secondary households, and the prevalence of infection) can all be efficiently evaluated by solving systems of linear equations (for example, using the backslash operator in MATLAB). In particular, these quantities are based upon the expectation, or distribution, of a path integral of a Markov chain [41–43]; we therefore present the necessary aspects of this theory, for the most part taken from Pollett and Stefanov [42].

### Expectation and distribution of path integrals for Markov chains

Let  $(X(t), t \geq 0)$  be a continuous-time Markov chain taking values on a finite subset of the non-negative integers  $S = A \cup C$ , where  $C$  is a set of transient states and  $A$  is a set of absorbing states, so the chain is absorbed almost-surely in finite time. (For our household dynamics, when  $\varepsilon = 0$ ,  $A$  refers to when the entire household is susceptible and  $C$  is the set of all other (transient) states for a household). Now consider a function  $f(\cdot) : S \rightarrow [0, \infty)$  with  $f(k) = 0$  for  $k \in A$ ;  $f(j)$  may be thought of as a per-unit reward when in state  $j \in C$  or for the households model can be naturally used to count the number of infected individuals in a given household configuration. Now, consider the path integral

$$\Gamma = \int_0^\infty f(X(t)) dt.$$

This path integral is the total reward over the life of the process, and as such is a random variable. We now present systems of linear equations for evaluating the expected value and Laplace-Stieltjes transform of the distribution of  $\Gamma$ .

**Table 1.** Classical within-household *SIRS* model of epidemic dynamics.

Event	Transition	Rate
Internal Infection	$(S, I) \rightarrow (S-1, I+1)$	$q((S, I), (S-1, I+1)) = \beta SI / (N-1)$
Recovery	$(S, I) \rightarrow (S, I-1)$	$q((S, I), (S, I-1)) = \gamma I$
Waning Immunity	$(S, I) \rightarrow (S+1, I)$	$q((S, I), (S+1, I)) = \mu(N-S-I)$
External Infection	$(S, I) \rightarrow (S-1, I+1)$	$q((S, I), (S-1, I+1)) = \varepsilon S$

doi:10.1371/journal.pone.0009666.t001

The behaviour of the continuous-time Markov chain can be defined by fixed transition rates between states which we formulate into a matrix  $Q = (q(i,j), i,j \in \mathcal{S})$ , with  $q(i,j)$  representing the rate of transition from state  $i$  to state  $j$ , for  $j \neq i$ , and  $q(i,i) = -q(i)$ , where  $q(i) = \sum_{j \neq i} q(i,j)$  ( $< \infty$ ), is the total rate at which the process leaves state  $i$ .

The expected value of  $\Gamma$  conditional on starting the process in state  $i$ , namely  $E[\Gamma | X(0) = i] = e_i$ , may be determined by the solution of the system of linear equations (Proposition 2 [42]):

$$\sum_{j \in \mathcal{C}} q(i,j)e_j + f(i) = 0, \quad i \in \mathcal{C}. \quad (1)$$

Furthermore, letting  $y_i(s) = E[e^{-s\Gamma} | X(0) = i]$  (the Laplace-Stieltjes transform of the distribution of  $\Gamma$ ) and noting  $y_i(s) = 1$  for  $i \in \mathcal{A}$ , we have [42, Proposition 1]: For each  $s > 0$ ,  $y_i(s)$  is the solution of the system of linear equations:

$$\sum_{j \in \mathcal{S}} q(i,j)y_j(s) = sf(i)y_i(s), \quad i \in \mathcal{C}. \quad (2)$$

Using these two sets of linear equations it is possible to efficiently evaluate a number of quantities concerning the early dynamics of infection within structured households models.

### Households of size one

To illustrate the power of this methodology, we consider the dynamics of households of size one with *SIR* dynamics for which many of the key epidemiological quantities are already known. (This can also be realised by setting the within-household transmission rate,  $\beta$ , to zero, but the computation is more complex.) Assuming that the number of households is infinite, then an initial infectious individual creates a new infectious household (individual) at rate  $\alpha$  over the course of their infectious lifetime, and moves from the infected to recovered state at rate  $\gamma$ . The Markov process for within-household dynamics can be formulated by letting  $X(t) = 1$  if the individual is infectious, and  $X(t) = 0$  if it has recovered; we therefore have  $q(1,0) = \gamma$  (and thus  $q(1,1) = -\gamma$ ).

We are naturally interested in onward transmission from this household, which means that we wish to consider  $f(X(t)) = \alpha X(t)$ . Then, from (1), the expected number of secondary households infected by this infectious individual over its lifetime,  $R_*$  (in this case equivalent to  $R_0$ ) is the solution  $e_1$  to:  $-\gamma e_1 = -\alpha$ . Hence  $R_* = e_1 = \alpha/\gamma$  as expected from fundamental results for the *SIR* model.

Now, for future use, consider the distribution of the path integral. We have from (2):  $\gamma - \gamma y_1(s) = s\alpha y_1(s)$ , and thus  $y_1(s) = \gamma/(s\alpha + \gamma)$ , and upon inverting this Laplace transform we find that  $\Gamma$  is exponentially distributed with mean  $\alpha/\gamma$  – that is,  $\Gamma$  has a probability density function  $\rho_\Gamma(\lambda) = (\gamma/\alpha)\exp(-\lambda\gamma/\alpha)$ . Whilst we will not be interested in this distribution directly *per se*, it does allow us to evaluate the full (discrete) distribution of secondary infections: the offspring distribution is Poisson with random mean, and the probability density function of the mean is  $\rho_\Gamma(\lambda)$  [18,44]. Thus, the probability mass function of secondary infections,  $g(m)$  (the probability that an infected individual generates  $m$  secondary cases), for any integer  $m$  is given by

$$\begin{aligned} g(m) &= \int_0^\infty \rho_\Gamma(\lambda) \frac{e^{-\lambda} \lambda^m}{m!} d\lambda \\ &= \frac{1}{m!} E[e^{-s\Gamma} \Gamma^m]_{s=1}. \end{aligned}$$

We note here that we may substitute  $\rho_\Gamma(\lambda)$  and integrate to obtain

$$g(m) = \frac{\alpha^m \gamma}{(\alpha + \gamma)^{m+1}},$$

which is the geometric distribution with parameter  $\alpha/(\alpha + \gamma)$ , as expected from the basic dynamics. However, by noting the close resemblance to the Laplace-Stieltjes transform introduced earlier, we develop an alternative procedure which is in general more efficient. Consideration of the expectation component gives

$$\begin{aligned} E[e^{-s\Gamma} \Gamma^m] &= (-1)^m \frac{d^m}{ds^m} E[e^{-s\Gamma}] \\ &= (-1)^m \frac{d^m}{ds^m} y(s). \end{aligned}$$

Hence,  $g(0)$  may be determined as the solution of the system of linear equations (2), as  $g(0) = y_1(1)$ . Then, differentiating the system (2), we have

$$\sum_{j \in \mathcal{C}} q(i,j)y_j^m(s) = mf(i)y_i^{m-1}(s) + sf(i)y_i^m(s), \quad i \in \mathcal{C}, \quad (3)$$

where  $y_i^m(s)$  denotes the  $m$ th derivative of  $y_i(s)$ , allowing us to recursively evaluate the distribution of secondary infections from

$$g(m) = (-1)^m \frac{y_1^m(1)}{m!}$$

(we may stop the recursive evaluation once the cumulative probability mass is close to 1). For the case of households of size one, we readily arrive at the geometric distribution with parameter  $\alpha/(\alpha + \gamma)$ , as evaluated earlier by direct integration.

Finally, we consider the early growth rate  $r$  of the epidemic (equivalent to the *Malthusian parameter* of the branching process approximation of Ball *et al.* [21]), defined as the solution to the equation

$$E\left[\int_0^\infty \alpha X(t)e^{-rt} dt\right] = 1.$$

We note that this is again the expectation of a path integral, but with exponential discounting at rate  $r(>0)$ ; as presented in Norris [41], and easily seen, this is equivalent to the path integral of the original process modified so that from each state  $i \in \mathcal{C}$  we add a rate  $r$  of jumping to the absorbing set  $\mathcal{A}$ . Thus, for households of size one:  $-(\gamma + r)e_1 = -\alpha$ , and upon solving  $e_1 = 1$  for  $r$  we determine  $r = \alpha - \gamma$ , a well-known result from the literature.

### A population of households

We now turn our attention to evaluating each of the desired quantities for households of arbitrary, but homogeneous, size. Populations of mixed household sizes are achievable through the same basic methodology, but the ensuing behaviour is more difficult to visualise and the relationships to the distribution of household sizes is naturally more complex. When dealing with household dynamics we initially consider a Markov chain model of an epidemic occurring within the initially infected household, effectively setting  $\varepsilon = 0$ , and determine the subsequent dynamics by considering the rate at which subsequent households become infected – mirroring the mechanisms used for households of size one. For simplicity, from now on we rescale time such that the

recovery rate  $\gamma = 1$ ; effectively our new unit of time is the average infectious period.

**(i) Household basic reproduction number,  $R_*$ .** The household basic reproduction number  $R_*$  is calculated as the expected number of secondary households infected by the primary household, which naturally scales with the amount of infection within the primary household. We define  $I(X(t))$  as the function of the underlying household Markov chain that gives the number of infected individuals within the household at time  $t$ . The household basic reproduction number is then given by

$$R_* = E \left[ \int_0^\infty \alpha I(X(t)) dt \right],$$

where  $X(0)$  refers to a household with a single infection and the remaining members susceptible [21,24,36]. This is calculated as described above, solving the system of linear equations (1). Evaluating  $R_*$  with  $\alpha=1$  gives us the critical value  $\alpha_c = 1/R_*(\alpha=1)$ , at which  $R_*=1$  and is therefore the invasion threshold.

**(ii) Early growth rate.** The early growth rate  $r$  obeys

$$E \left[ \int_0^\infty \alpha I(X(t)) e^{-rt} dt \right] = 1.$$

This can be derived either from consideration of the branching process approximations [21], or a survivor model [33], and can be evaluated using exponential discounting as outlined above [41]. Since  $r \in [-1, R_* - 1]$  [45] (here we only consider epidemics which invade, and thus  $r > 0$ ) and the expectation monotonically decreases with  $r$ , a unique solution exists and can be found numerically using, for example, interval bisection or MATLAB's `fzero` routine.

**(iii) Household offspring distribution.** The household offspring distribution during the early stages of the epidemic is again Poisson with random mean:

$$\begin{aligned} g(m) &= \int_0^\infty \rho_\Lambda(\lambda) \frac{e^{-\lambda} \lambda^m}{m!} d\lambda \\ &= \frac{(-1)^m}{m!} y_{IC}^m(1), \end{aligned} \quad (4)$$

for integer  $m$  and where  $\rho_\Lambda(\lambda)$  is the probability density function of the stochastic variable

$$\Lambda = \int_0^\infty \alpha I(X(t)) dt,$$

$y(s) = y^0(s)$  is the Laplace transform of the distribution of  $\Lambda$ , and  $y^m(s)$  is the  $m$ -th derivative of this transform with respect to  $s$ . As with the two above quantities, we are interested in the entry corresponding to the initial condition (IC)  $I = 1, S = N - 1$ . We can calculate the offspring distribution (4) by solving systems of linear equations (2,3), as outlined earlier for the case of households of size one.

**(iv) Endemic prevalence.** To evaluate the endemic prevalence, we could evaluate the full transient dynamics of the process, as explained in the next section, and evaluate the endemic prevalence via convergence to equilibrium of these dynamics. However, it is possible to develop a more efficient procedure, which does not rely on numerically solving the differential equations.

We exploit the fact that at equilibrium the rate of import of infection into a household,  $\varepsilon$ , must be equal to the rate of export of infection  $\alpha J^*$ . Throughout this paper, we use  $J = I/N$  to represent the proportion of individuals infected. Thus, our starting point for determining the endemic proportion of infection  $J^*$  in the population is by considering the dynamics within a single household, given by our within-household Markov chain model detailed in Table 1, with constant (and not yet self-consistent) external force of infection  $\varepsilon > 0$ . Using an eigenvalue-vector routine, for example the MATLAB function `eig`, we find the eigenvector corresponding to the zero eigenvalue; normalised to sum to 1, this eigenvalue  $\pi_\varepsilon = X_\varepsilon(\infty)$  is the stationary (equilibrium) distribution of the within-household epidemic with import at rate  $\varepsilon$ . We then search for the  $\varepsilon$  such that  $\varepsilon = \alpha I(X_\varepsilon(\infty))/N$ , which can be done with considerable efficiency given that  $I(X_\varepsilon(\infty))$  increases monotonically with  $\varepsilon$ . The endemic prevalence  $J^* = I(X_\varepsilon(\infty))/N$ .

**(v) Transient dynamics.** To model the full dynamics of the system we need to both include the external rate of infection  $\varepsilon$  and to allow it to dynamically vary in a self-consistent manner [46]. Let  $H(t)$  be the vector of proportions of households in each possible disease configuration at time  $t$ ; its dynamics are described by the coupled ODEs:

$$\frac{dH}{dt} = HQ_{\{\varepsilon = \alpha I(H(t))/N\}} \quad (5)$$

where  $Q$  is the household transition matrix together with the dynamically varying external force of infection. We note that this equation may also be motivated by using results of Kurtz [47], considering the proportion of household types in the limit as the number of households tends to infinity. Due to the fact that  $\varepsilon$  varies during the transient dynamics methodologies based on taking the exponential of matrices [48] are not applicable. The full dynamical system (5) is therefore solved numerically using standard Runge-Kutta methods, for example as implemented in MATLAB's `ode45`.

**Additional realism.** As noted earlier, we also consider the effect of gamma-distributed infectious periods on several of the above quantities. This is achieved by using the commonly-termed *method of stages*: the infectious class is decomposed into several subclasses, with an identical rate of transition between these classes chosen to retain the same expected infectious period. Such an approach maintains the Markov property of the underlying model and thus the applicability of the methodology outlined above (see for example [39] and [40]).

When comparing results of gamma-distributed to the traditional exponentially-distributed periods, we consider two cases: i) in the first we hold the transmission rate parameter  $\beta$  constant; ii) whilst in the second we hold the probability of (initial) transmission,  $p$ , constant. The latter is evaluated as follows: Consider a household with one individual initially infected and all other members susceptible to infection, with transmission rate  $\beta$ , mean infectious period 1, and gamma-distributed recovery time of order  $M$  (referring to dividing the infectious period into  $M$  sub-classes,  $M = 1$  is exponentially distributed and the limit  $M \rightarrow \infty$  gives a constant time to recovery). The probability that the initial infective infects at least one other individual before recovering is

$$p(\beta, M) = \frac{(\beta + M)^M - M^M}{(\beta + M)^M}.$$

Note,  $p(\beta, 1) = \beta/(\beta + 1)$  and  $p(\beta, \infty) = 1 - e^{-\beta}$ . We consider holding  $p$  constant (via changing  $\beta$ , to  $\beta'$ ) as we change  $M$  because this is similar to holding constant the *secondary attack rate*

(we discuss this further in the Supporting Information File S1), a quantity that is often measured during statistical analysis of household data [49].

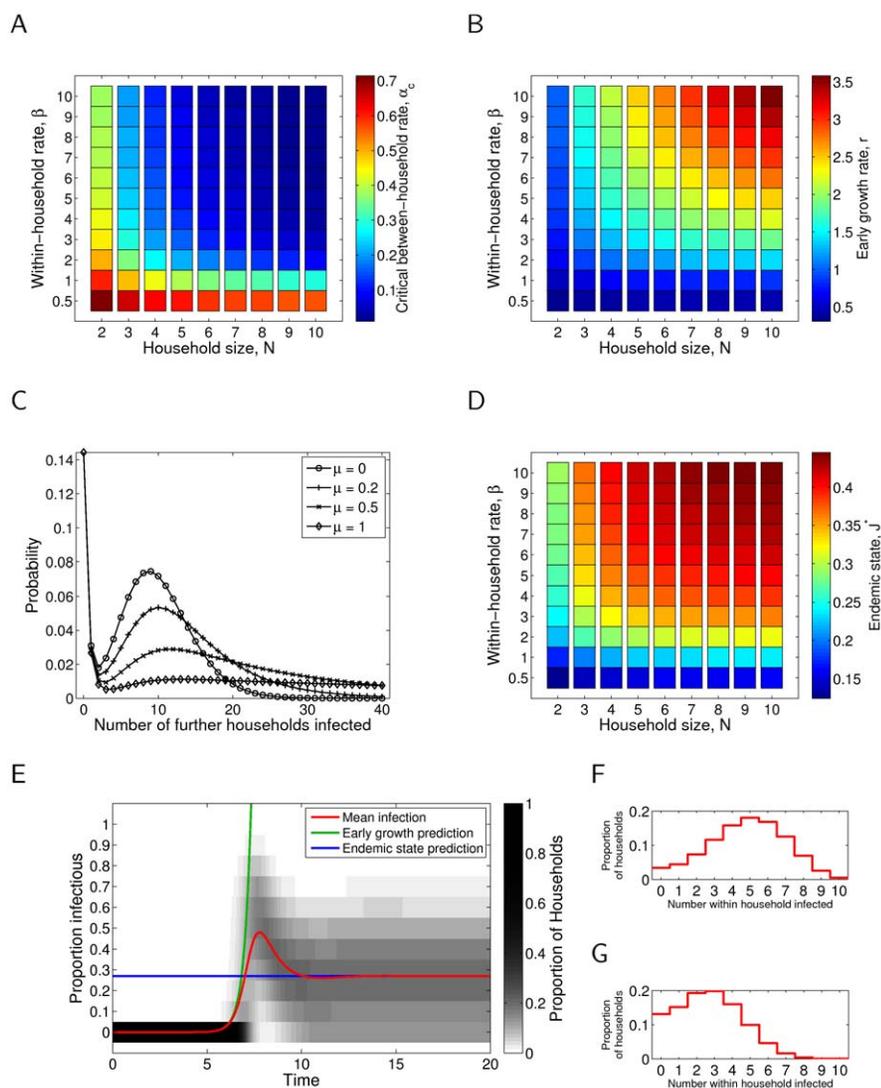
## Results

Figure 1 shows the critical level of between-household transmission,  $\alpha_c$ , required to sustain an epidemic (Plot 1A); the early growth rate  $r$  (Plot 1B); the offspring distribution (Plot 1C); the endemic prevalence proportion  $J^*$  (Plot 1D); and the transient dynamics of infection (Plot 1E; see 1F and 1G also), all for the case of exponentially distributed infectious period ( $M = 1$ ). The precise parameter values are provided in the Figure caption; other rate parameter values are presented in the Supporting Information File S1.

Examining the critical level of between-household transmission (Plot 1A) indicates that two factors contribute to the success of an

infection. For large household sizes, the main determinant of epidemic success is whether within-household transmission (governed by  $\beta$ ) can produce an epidemic within the household – if it can, then the final size within households is relatively large and so relatively small levels of between-household transmission,  $\alpha$ , can sustain an epidemic. For smaller households of size  $N = 2$  and  $N = 3$ , this is not seen, and appreciable between- and within-household transmission is always necessary to sustain an epidemic.

The endemic prevalence of infection (when between-household transmission rate  $\alpha = 1$ ; Plot 1D) is again predominately determined by within-household transmission,  $\beta$ , for larger households, with household size,  $N$ , only having a major impact when it is below size 4. However, unlike the critical level of transmission (Plot 1A) and early growth rate (Plot 1C), varying the rate of waning immunity,  $\mu$ , has a significant effect in terms of absolute prevalence  $J^*N$ , which reduces as the rate of loss of immunity  $\mu$



**Figure 1. Epidemiological quantities as defined in the main text for the exponential ( $M = 1$ ) infectious period distribution.** (A) Critical transmission  $\alpha_c$ , with rate of waning immunity  $\mu = 1$ . (B) Early growth rate  $r$ , with rate of waning immunity  $\mu = 1$  and between-household transmission parameter  $\alpha = 1$ . (C) Offspring distribution, with within-household transmission rate parameter  $\beta = 6$ , house size  $N = 10$  and between-household transmission parameter  $\alpha = 1$ . (D) Endemic infection  $J^*$ , with rate of waning immunity  $\mu = 1$  and between-household transmission parameter  $\alpha = 1$ . (E) Transient dynamics, with within-household transmission rate parameter  $\beta = 6$ , house size  $N = 10$ , rate of waning immunity  $\mu = 1$  and between-household transmission parameter  $\alpha = 1$ . (F) As in E at peak prevalence. (G) As in E at endemic prevalence. doi:10.1371/journal.pone.0009666.g001

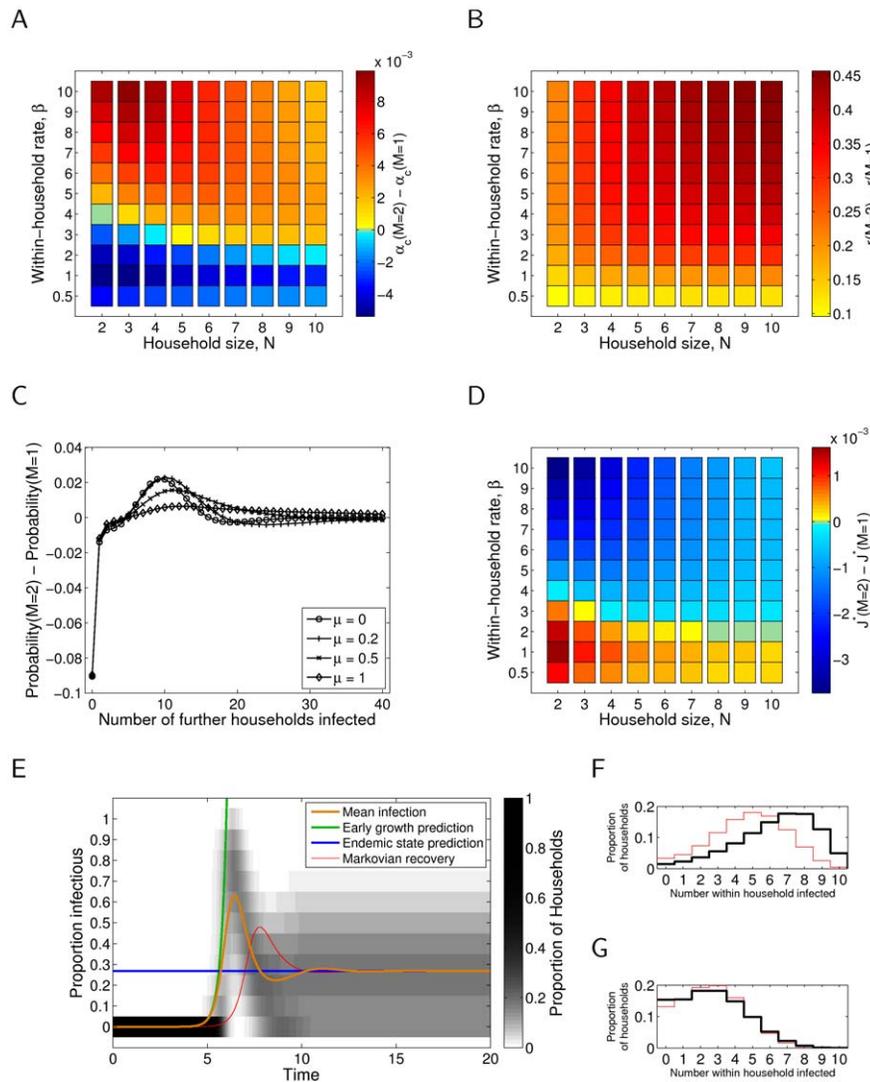
is reduced, and vanishes in the absence of waning immunity ( $\mu=0$ ); see Supporting Information File S1.

The early growth rate  $r$  (with between-household transmission  $\alpha=1$ ; Plot 1B) varies with both within-household transmission,  $\beta$ , and household size,  $N$ , and is only weakly affected by waning immunity,  $\mu$  (see Supporting Information File S1). We observe that the critical between-household transmission rate (Plot 1A) and the endemic prevalence of infection (Plot 1D) show far greater saturation with both household size,  $N$ , and within household transmission rate,  $\beta$ , compared to the early growth rate (Plot 1B). This is because the critical transmission rate,  $\alpha_c$ , and the endemic prevalence of infection,  $J^*$ , depend on the number of cases produced over one generation (which rapidly saturates as susceptibles within the household get infected), whereas the early growth rate,  $r$ , depends on the instantaneous transmission rate

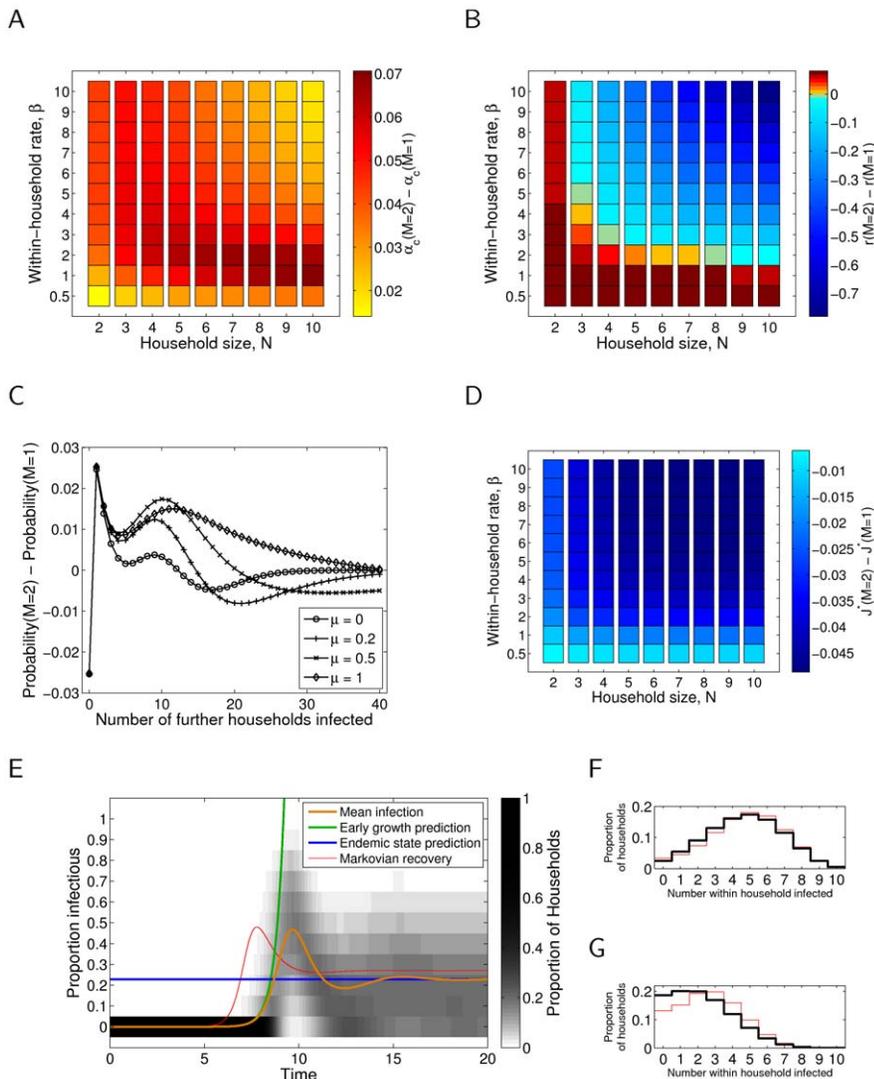
from infected individuals (which is less influenced by households reaching saturation).

The offspring distribution (Plot 1C) shows a significant probability that a newly infected household will fail to infect any further households; often because the infection fails to spread within the household. This failure probability is relatively unaffected by the waning immunity rate,  $\mu$ ; whereas increasing  $\mu$  leads to an increased probability of generating very large numbers of secondary cases. The bimodality of these distributions is a qualitative difference from distributions considered in detail at the individual level [50,51] and therefore can be expected to lead to very different stochastic invasion and persistence properties.

The complete model dynamics at a given parameter set (Plot 1E) demonstrates that our methods for calculating early growth (green line) and the endemic state (blue line) are sound, and also that the



**Figure 2. Epidemiological quantities as defined in the main text for the gamma ( $M > 1$ ) infectious period distribution, compared to the exponential ( $M = 1$ ) results, where within household transmission rate parameter,  $\beta$ , is held constant.** (A) Critical transmission difference  $\Delta\alpha_c$ , with rate of waning immunity  $\mu=1$ ;  $M=2$ . (B) Early growth rate difference  $\Delta r$ , with rate of waning immunity  $\mu=1$  and between-household transmission parameter  $\alpha=1$ ;  $M=2$ . (C) Offspring distribution difference, with within-household transmission rate parameter  $\beta=6$ , house size  $N=10$  and between-household transmission parameter  $\alpha=1$ ;  $M=2$ . (D) Endemic infection difference  $\Delta J^*$ , with rate of waning immunity  $\mu=1$  and between-household transmission parameter  $\alpha=1$ ;  $M=2$ . (E) Transient dynamics, with within-household transmission rate parameter  $\beta=6$ , house size  $N=10$ , rate of waning immunity  $\mu=1$  and between-household transmission parameter  $\alpha=1$ ;  $M=3$ ; exponential ( $M=1$ ) results are shown as a thin red line for comparison. (F) As in E at peak prevalence. (G) As in E at endemic prevalence. doi:10.1371/journal.pone.0009666.g002



**Figure 3. Epidemiological quantities as defined in the main text for the gamma ( $M > 1$ ) infectious period distribution, compared to the exponential ( $M = 1$ ) results, where probability of transmission,  $p$ , is held constant.** (A) Critical transmission difference  $\Delta\alpha_c$ , with rate of waning immunity  $\mu = 1$ ;  $M = 2$ . (B) Early growth rate difference  $\Delta r$ , with rate of waning immunity  $\mu = 1$  and between-household transmission parameter  $\alpha = 1$ ;  $M = 2$ . (C) Offspring distribution difference, with within-household transmission rate parameter  $\beta' \approx 3.29$  ( $\beta = 6$ ), house size  $N = 10$  and between-household transmission parameter  $\alpha = 1$ ;  $M = 2$ . (D) Endemic infection difference  $\Delta J^*$ , with rate of waning immunity  $\mu = 1$  and between-household transmission parameter  $\alpha = 1$ ;  $M = 2$ . (E) Transient dynamics, with within-household transmission rate parameter  $\beta' \approx 2.74$  ( $\beta = 6$ ), house size  $N = 10$ , rate of waning immunity  $\mu = 1$  and between-household transmission parameter  $\alpha = 1$ ;  $M = 3$ ; exponential ( $M = 1$ ) results are shown as a thin red line for comparison. (F) As in E at peak prevalence. (G) As in E at endemic prevalence. doi:10.1371/journal.pone.0009666.g003

proportion of households with a given prevalence assumes a unimodal distribution around the mean, with significant variance.

Figures 2 and 3 essentially repeat the evaluation of each of the epidemiologically relevant quantities for the case of gamma distributed infectious period of order  $M > 1$ , with  $\beta$  and  $p$  held constant, respectively. In panels (A), (B), (C) and (D) the change in the quantity (respectively,  $\alpha_c$ ,  $r$ , probability of offspring number, and  $J^*$ ) with respect to the assumption of an exponential infectious period has been presented; likewise, in (E), the mean prevalence curve and, in (F) and (G), the full distribution of prevalence, under the assumption of an exponential infectious period have been superimposed for reference.

With respect to the critical transmission rate,  $\alpha_c$ , the early growth rate,  $r$ , and the endemic prevalence of infection,  $J^*$ , it can be seen that both the magnitude and direction of change can differ

depending upon whether the transmission parameter  $\beta$  or the probability of transmission,  $p$ , is held constant. With respect to these quantities, holding  $p$  constant generally results in more significant changes. Furthermore, in general, a higher level of between-household transmission,  $\alpha$ , is required to sustain an epidemic, and thus the endemic prevalence of infection,  $J^*$ , with  $\alpha = 1$  is generally reduced. The early growth rate increases if  $\beta$  is held constant (Plot 2B) and generally decreases if  $p$  is held constant (Plot 3B).

With respect to the offspring distributions (C), once again the incorporation of gamma-distributed infectious period, and choice of what is held constant between epidemics, has a significant impact. In both cases the probability of no secondary households infected decreases, and by a substantial margin in the case of  $\beta$  held constant. In the  $\beta$  constant case the probability of a small number of secondary infections also decreases, whilst in the  $p$  constant case this

is not seen. In both cases there is a decrease in the tail of the distribution, corresponding to reduced probability of a large number of secondary infections, and the major increase in probability mass occurs in the vicinity of the second peak of the exponential offspring distribution case. Despite these changes, for this parameter set, the bimodal feature of these distributions remains.

Finally, we consider the influence of gamma-distributed infectious period, and what is held constant, by studying the full dynamics of infection. Holding  $\beta$  constant (Plot 2E) results in an earlier epidemic with a larger peak infection when compared to the exponential case, whilst holding  $p$  constant (Plot 3E) results in an epidemic with similar, but slightly reduced, peak incidence but slower take-off and hence delayed peak. Also interestingly, the incorporation of gamma-distributed infectious period results in a slight oscillatory approach to endemicity, with mean infection dropping below the endemic prevalence following the peak, and slightly overshooting the endemic level again before converging to equilibrium.

## Discussion

In this paper, we have focused on appropriate numerical methods for efficient calculation of epidemiologically relevant quantities in households models obeying an *SIRS* disease paradigm. Solutions of the kind we have found provide a useful bridge between formal work [36] and individual-based simulation [24], allowing us to study the effects of finite size, stochasticity, and infectious and recovered period distributions at the household level while still in the infinite-size limit at the population level. This has allowed us to quantify some basic behaviours of household structured dynamics; for example the impacts of varying within-household and between-household transmission rates and the role of waning immunity.

In essence the behaviour of the households model can be explained by two different processes. Firstly, compared to purely individual-based transmission (at rate  $\alpha$ ) the action of within-household transmission (at rate  $\beta$ ) is to amplify the infection; so households act as amplifiers for the general transmission process. Secondly, the clustered network structure within a household leads to rapid depletion of the locally susceptible population and hence saturation of amplification effect (and other household properties) as the within-household transmission rate increases. Using our methods, a full sweep of parameter space (as provided in Supporting Information File S1) requires only a few seconds of desktop processor time and so comprehensive sensitivity analysis around a region of parameter space relevant to a given applied problem is also readily obtained. In addition, such methods are able to rapidly calculate likelihood values for any given observations, leading to efficient methods of parameter estimation from household-structured data.

Our results additionally provide one of the first explicit studies of the impact of gamma-distributed infectious periods on dynamics at the household level. Biologically significant effects arise from this change in infectious period distribution, with both the magnitude and direction of these effects varying depending on which parameters are held constant when comparing models. This link between which quantities are observed and therefore which model parameters are held constant, is likely to have significant

impact during the early stages of a disease outbreak in determining an appropriate response.

The methodology used to consider gamma-distributed infectious periods can be straightforwardly extended to include latent and prodromal stages of infection. For homogeneous mixed models, the most common such addition is an ‘exposed’ class through which infected individuals pass before becoming fully infectious, leading to the standard *SEIR* model. This does not alter threshold and final size behaviour compared to the *SIR* model, but can cause a highly significant reduction of the early growth rate and modification of other features of the transient dynamics, which we would also expect to see in household models.

We also believe that the epidemiological consequences of the shape of the household offspring distribution warrant further consideration, as has been done for individual-level offspring distributions [50,51]. The ability to construct this distribution through the solution of sets of linear equations offers the possibility of deriving such a distribution, and therefore a greater understanding of stochastic invasion and persistence, for a range of more complex disease natural histories.

Methodologically, we hope that the reduction of many problems in household epidemic theory to solving a set of linear equations through the theory of path integrals for Markov chains will be of significant use, and have made MATLAB code available to encourage this in the Supporting Information File S2. It is important to note that the consideration of distributions of household sizes is simply done within our framework, but is not included in this work since its main impact is known, from theoretical work, to be on control strategy rather than dynamics. While it could be argued that the methodologies presented here simply allow us to produce the same results more quickly than by using direct integration, numerically efficient algorithms can open up problems to analysis that are currently unsolvable. One example of this would be to consider epidemiological dynamics of sub-populations much larger than households, either ecologically-motivated or as a simplification of large-scale structured models of human and animal disease.

## Supporting Information

**File S1** Calculation of disease dynamics in a population of households. In this Supporting Information File S1, we include additional technical discussion, and supplementary figures.

Found at: doi:10.1371/journal.pone.0009666.s001 (0.08 MB PDF)

**File S2** Calculation of disease dynamics in a population of households. In this Supporting Information File S2 we provide MATLAB code.

Found at: doi:10.1371/journal.pone.0009666.s002 (0.01 MB ZIP)

## Author Contributions

Conceived and designed the experiments: JR TH MJK. Performed the experiments: JR TH MJK. Analyzed the data: JR TH MJK. Contributed reagents/materials/analysis tools: JR TH MJK. Wrote the paper: JR TH MJK.

## References

1. Kermack W, McKendrick A (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A* 115: 700–721.
2. Anderson R, May R (1992) *Infectious diseases of humans*. Oxford: Oxford University Press.
3. Keeling M, Rohani P (2008) *Modeling infectious diseases in humans and animals*. Princeton: Princeton University Press.
4. Bartlett M (1956) Deterministic and stochastic models for recurrent epidemics. *Proc 3rd Berkeley Symp Math Statist Probab* 4: 81–109.
5. Bailey N (1957) *The mathematical theory of epidemics*. London: Griffin.
6. Andersson H, Britton T (2000) *Stochastic Epidemic Models and Their Statistical Analysis*. Berlin: Springer Lectures Notes in Statistics, Vol. 151, Springer.
7. Diekmann O, Heesterbeek J (2000) *Mathematical Epidemiology of Infectious Diseases Model Building, Analysis and Interpretation*. England: Wiley.
8. Altmann M (1995) Susceptible-infectious-recovered epidemic models with dynamic partnerships. *J Math Biol* 33: 661–675.

9. Keeling M, Rand D, Morris A (1997) Correlation models for childhood epidemics. *Proc R Soc Lond B* 264: 1149–1156.
10. Watts D, Strogatz S (1998) Collective dynamics of “small-world”.
11. Keeling M (1999) The effects of local spatial structure on epidemiological invasions. *Proc R Soc Lond B* 266: 859–867.
12. Potterat J, Rothenberg R, Muth S (1999) Network structural dynamics and infectious disease propagation. *Int J STD AIDS* 10: 182–185.
13. Klondahl A (2001) Networks and pathogens. *Sex Transm Dis* 28: 25–28.
14. Halloran M, Longini I, Nizam A, Yang Y (2002) Containing bioterrorist smallpox. *Science* 298: 1428–1432.
15. Keeling M, Eames K (2005) Networks and epidemic models. *J Roy Soc Interface* 2: 295–307.
16. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: Structure and dynamics. *Physics Reports* 424: 175–308.
17. Dangerfield C, Ross J, Keeling M (2009) Integrating stochasticity and network structure into an epidemic model. *J Roy Soc Interface* 6: 761–774.
18. Ball F (1986) A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemics models. *Adv App Prob* 18: 289–310.
19. Addy C, Longini Jr. I, Haber M (1991) A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* 47: 961–974.
20. Becker N, Dietz K (1995) The effect of household distribution on transmission and control of highly infectious diseases. *Math Biosci* 127: 207–219.
21. Ball F, Mollison D, Scalia-Tomba G (1997) Epidemics with two levels of mixing. *Ann App Prob* 7: 46–89.
22. Ball F, Neal P (2002) A general model for stochastic sir epidemics with two levels of mixing. *Math Biosci* 180: 73–102.
23. Cross P, Lloyd-Smith J, Johnson P, Getz W (2005) Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecol Lett* 8: 587–595.
24. Cross P, Johnson P, Lloyd-Smith J, Getz W (2007) Utility of  $\tau_0$  as a predictor of disease invasion in structured populations. *J Roy Soc Interface* 4: 315–324.
25. House T, Keeling M (2008) Deterministic epidemic models with explicit household structure. *Math Biosci* 213: 29–39.
26. Ball F, Sirl D, Trapman P (2009) Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv App Prob* 41: 765–796.
27. Ferguson N, Cummings D, Fraser C, Cajka J, Cooley P, et al. (2006) Strategies for mitigating an influenza pandemic. *Nature* 442: 448–452.
28. Pellis L, Ferguson N, Fraser C (2009) Threshold parameters for a model of epidemic spread among households and workplaces. *J R Soc Interface* 40: 979–987.
29. Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. *PNAS* 101: 15124–15129.
30. Kao R, Danon L, Green D, Kiss I (2006) Demographic structure and pathogen dynamics on the network of livestock movements in great britain. *Proc R Soc Lond B* 273: 1999–2007.
31. Robinson S, Everett M, Christley R (2007) Recent network evolution increases the potential for large epidemics in the british cattle population. *JRS Interface* 4: 669–674.
32. Becker N, Utev S (1998) The effect of community structure on the immunity coverage required to prevent epidemics. *Math Biosci* 147: 23–39.
33. Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* 2: e758.
34. Wu J, Riley S, Fraser C, Leung G (2006) Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med* 3: e361.
35. House T, Keeling M (2008) Household structure and infectious disease transmission. *Epidemiol Infect* 137: 654–661.
36. Ball F (1999) Stochastic and deterministic models for sis epidemics among a population partitioned into households. *Math Biosci* 156: 41–67.
37. Ball F, Lyne O (2001) Stochastic multitype sir epidemics among a population partitioned into households. *Adv App Prob* 33: 99–123.
38. Koopman J, Chick S, Simon C, Riolo C, Jacquez G (2002) Stochastic effects on endemic infection levels of disseminating versus local contacts. *Math Biosci* 180: 49–71.
39. Keeling M, Grenfell B (2002) Understanding the persistence of measles: reconciling theory, simulation and observation. *Proc R Soc Lond B* 269: 335–343.
40. Ross J, Pagedam D, Pollett P (2009) On parameter estimation in population models ii: multi-dimensional processes and transient dynamics. *Theor Pop Biol* 75: 123–132.
41. Norris J (1997) *Markov chains*. Cambridge: Cambridge University Press.
42. Pollett P, Stefanov V (2002) Path integrals for continuous-time markov chains. *J App Prob* 39: 901–904.
43. Pollett P (2003) Integrals for continuous-time markov chains. *Math Biosci* 182: 213–225.
44. Daniels H (1967) The distribution of the total size of an epidemic. *Proc 5th Berkeley Symp Math Statist Prob* 4: 281–293.
45. Goldstein E, Paur K, Fraser C, Kenah E, Wallinga J, et al. (2009) Reproductive numbers, epidemic spread and control in a community of households. *Math Biosci* 221: 11–25.
46. Ghoshal G, Sander L, Sokolov I (2004) Sis epidemics with household structure: the self-consistent field method. *Math Biosci* 190: 71–85.
47. Kurtz T (1970) Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability* 7: 49–58.
48. Keeling M, Ross J (2008) On methods for studying stochastic disease dynamics. *J Roy Soc Interface* 5: 171–181.
49. Longini Jr. I, Koopman J, Monto A, Fox J (1982) Estimating household and community transmission parameters for influenza. *Am J Epi* 115: 736–751.
50. Lloyd-Smith J, Schreiber S, Kopp P, Getz W (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* 438: 355–359.
51. James A, Pitchford J, Plank M (2007) An event-based model of superspreading in epidemics. *Proc R Soc Lond B* 274: 741–747.