Journal of Progressive Research in Mathematics
Vol 18,3 - 2021 — E-ISSN: 2395-0218
http://www.scitecresearch.com/journals/index.php/jprm

Stochastic SIR Household Epidemic Model with Misspecification

Umar ,M.A

[1] Department of Mathematics, Nasarawa State University, Keffi Nigeria
Email: abdulmallam@yahoo.com

Received: June 20, 2021; Accepted: July 24, 2021; Published: August 9, 2021

Abstract

The stochastic SIR household epidemic model is well discussed in [3], [4], [5] and also in [1] by assuming that the infection period distribution is known. Sometimes this may wrongly be assumed in the model estimation and hence the adequacy of the model fitness to the final size data is affected. We examined this problem using simulations with large population size and theoretical parameters in which the final size data is first simulated with $\exp(4.1)$ **infectious period distribution and estimated with** $\Gamma(2, 4.1/2)$ **infectious period distribution and vice versa. The estimates of the two dimensional models are further explored for a range of local and global infection rates with corresponding proportion infected and found to be biased and imprecise.**

Keywords:

Final size epidemic, Infectious period distribution, Maximum likelihood estimates, Misclassification probabilities.

1. Introduction

Sometimes the final size epidemic data may be wrongly estimated with an infectious period distribution different from that used in simulation. This is a misspecification problem that is sometimes encountered in modelling process.

In this work, we examined this problem using simulations with large population size, and choice of appropriate theoretical parameters which allows global infection. Our studies is primarily focused on misspecification of the infectious period distribution, as extensive studies on the model behaviors, its theoretical parameters and other properties have been provided by various researchers especially in [5, 6, 2], [5], [7], [8],[9], [10],[13], [18], [19] and more recently [22], whose work examined effects of minimum epidemic size and population sizes on a global epidemic in simulation of final size epidemic data and the behavior of the estimates when the theoretical parameters correspond $z = 0.1775$ and $z = 0.7298$ with minimum epidemic size of 1000 for various local and global infection rates. Also, the properties of the parameters and functions of the stochastic SIR household epidemic model are well discussed for ease of understanding of the model

behaviours and also allow comparison of results from these studies with those that employed same infectious periods distribution in their simulations and model estimation as in [5, 6, 2] and [22]

This work therefore examined the stochastic SIR household epidemic model with some their properties as found in [4], [11], [15], [14], [17],[20, 21], [16] and a host of other authors, its estimation in the face of misspecification of the infectious period distribution is also discussed.

2. Material and Methods

2.1. The model.

Discussion on the stochastic SIR household epidemic model can be found in [5, 6, 2], especially in [5], one of the pioneering research figure on this topic. Other extensions are provided in the works of [1], [4], [3], [12], which allows construction of likelihood function often referred to as approximation likelihood function as in [1], [22] and [5] hence enables model estimation based on the assumption that the infectious period distribution is often the same as the one used in simulation.

If the stochastic SIR household epidemic model is estimated using a different infectious period distribution from that used for the simulations, then how does this affect the precision of the estimates? This is a misspecification problem which may sometimes be taken for misclassification of the epidemic data. It is therefore necessary to study these scenarios using simulations in order to understand their effects on the estimates of the parameters. We do this with large population size and theoretical parameters, $\lambda_L = 0.1$, $\lambda_G = 0.29$ which give global infection in our simulations and hence enable us compare the estimates with those of our previous studies.

We simulate two dimensional model epidemic data with exp(4.1) infectious period distribution and estimated the model parameters with the Gamma$(2, 4.1/2)$ infectious period distribution. Plots of the estimates and tables of mean, standard deviation and root mean square errors are presented.

2.2. Numerical simulations of the Stochastic Household Epidemic.

In order to illustrate the threshold behaviour of SIR household epidemic model, we conducted 1000 simulations of a household epidemic for different values of the local and global infection rates, ,(λ_L, λ_G), with a modified version of the simhouses simulation package of Dr Owen Lyne, [1] household structure $[133, 189, 108, 106, 31] \times 50$.

Here, the entries represent number of households which size corresponds to its column. For example 133 is the number of households of size 1, 189 is the number of households of size 2, 108 is the number of households of size 3. The population is made of households of sizes 1 to 5 in which the number of households of each size is 50 times that of [1] and a population size of 70700. Also, we have assumed Gamma$(2, 2.05)$ infectious period distribution in [1] which has probability density function, $f_{T_I}(t) = c^2 t \exp(-ct)$, $c > 0$, where $c = 2/4.1$ and mean $E(T_I) = 4.1$ [1, 6].

Six pairs of parameter values, (λ_L, λ_G) are considered together with their corresponding threshold parameter in order to study the influence of the infection rates on the occurrence of a global epidemic in the simulation runs. Two columns of histograms of the number of individuals infected from the simulations are presented, with the one on the left having fixed global contact rate and varying local infection rates while those on the right hand side have fixed local infection rate and varying global infection rates.

Form the histograms of the number infected we see that the threshold behaviour exhibits the expected theoretical result such that when $R_* > 1$, then global epidemic occurs with probability $1 - p^a$, where $a = 1$ is the initial number of infectives. The bimodal behaviour of the histograms when $R_* > 1$ further clarify the occurrence a global epidemic in such cases. Thus, large epidemic only occurs when $R_* > 1$ in accordance with [5, 6], also given R_*, the precise values of λ_L and λ_G have little effect on either the number of people infected or the probability of large epidemic occurring.

Thus, the first two histograms at the top correspond to the case in which $R_* < 1$ and therefore global epidemic never occurred, while the remaining histograms are made of few cases in which a global epidemic

occur with bimodal behaviours and few cases in which there is no global epidemic.

In order to disallow the non global epidemic from occurring, we employed a minimum cut-off of the number infected between the epidemics using rejection sampling in which if the number infected in the simulation is less than the cut-off then it is rejected and a re-run is made. This is continued until the simulation run is completed.

3. Theoretical properties of the parameters and functions of the SIR household epidemic model.

In this chapter, we examined the theoretical properties of the parameters and functions of the stochastic SIR household epidemic model beginning with the mean final size of the household epidemic, the beta function for small and large local infection rates, the threshold parameter, the proportion of the initial susceptibles infected in a household epidemic

3.1. The mean final size of single household epidemic.

The mean final size of a single household epidemic is given in [5] and is defined as the average number of initial susceptibles that are ultimately infected, including the initial number of infectives, at the end of the disease outbreak expressed as

$$\mu_{n,a} = n + a - \sum_{k=0}^{n} \binom{n}{k} \beta_k \phi(\lambda_L k)^{n+a-k},$$

where n is the total number of susceptibles, a is the initial number of infectives at the beginning of disease outbreak, β_k are functions of λ_L and the infectious period distribution, obtained for $k \in \mathbb{Z}_+$ from the triangular equation in [4] as,

$$\sum_{i=0}^{k} \binom{k}{i} \beta_i \phi(\lambda_L i)^{k-i} = k, \quad k = 0, 1, 2, \ldots,$$

where, $\phi(\theta) = E(\exp(-\theta T_I))$, is the moment generating function of the infective period, T_I, and λ_L is the local contact rate. This can be expanded as,

$$\binom{k}{0} \beta_0 \phi(\lambda_L.0)^{k-0} + \binom{k}{1} \beta_1 \phi(\lambda_L.1)^{k-1} + \cdots + \binom{k}{k-1} \beta_{k-1} \phi(\lambda_L \cdot (k-1))^1 + \beta_k \phi(\lambda_L.k)^0 = k.$$

Observe that if $k = 0$, then $\beta_0 = 0$. Thus we can ignore the first term and express the equation as,

$$\binom{k}{1} \beta_1 \phi(\lambda_L.1)^{k-1} + \binom{k}{2} \beta_2 \phi(\lambda_L.2)^{k-2} + \cdots + \binom{k}{k-1} \beta_{k-1} \phi(\lambda_L \cdot (k-1))^1 + \beta_k = k.$$

We can also rearrange it as,

$$\beta_k = k - \sum_{i=1}^{k-1} \binom{k}{i} \beta_i \phi(\lambda_L i)^{k-i}. \tag{1}$$

3.2. Properties of β_k for small and large local infection rates.

If $\lambda_L \to 0$, then $\phi(\lambda_L) = E(\exp(-\lambda_L T_I)) \to 1$, $\forall\ T_I$ and equation 1 reduces to

$$\sum_{i=1}^{k} \binom{k}{i} \beta_i = k.$$

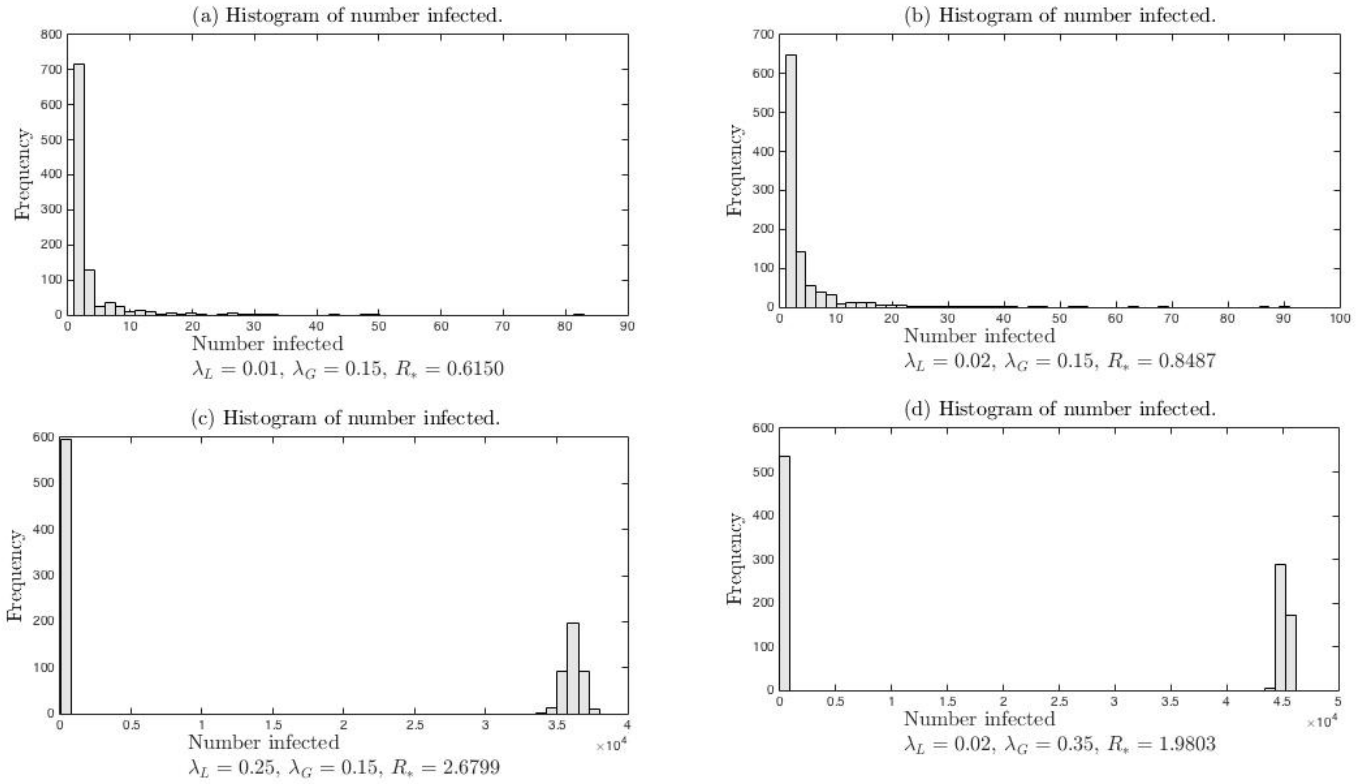

Figure 1: Histogram of 1000 simulations of household epidemic with Gamma$(2, 2.05)$ infectious period distribution, parameter estimates from [1] but fifty times its population size and minimum epidemic size of 1.

It follows that, if $\lambda_L \to 0$, β_k can be expressed as,

$$\beta_k = k - \sum_{i=1}^{k-1} \binom{k}{i} \beta_i.$$

Theorem 1. *If $\lambda_L = 0$, then $\beta_k = 0$, $\forall k \in \mathbb{Z}_+ - \{1\}$ and $\beta_1 = 1$ when $k = 1$.*

Proof. Using mathematical induction, we will show that $\beta_k = 0$, $\forall k \in \mathbb{Z}_+ - \{1\}$, whenever $\lambda_L = 0$.

From the arguments in equation (1), we know that $\beta_0 = 0$, when $k = 0$, also when $k = 1$, $\beta_1 = 1$, however when $k = 2$, then $\beta_2 = 0$.

Using mathematical induction, we want to show that $\beta_k = 0$, $\forall k \in \mathbf{Z}_+ - \{1\}$.

We assume the induction hypothesis holds for $\forall n \in \{2, \ldots, k\}$ and show that it also holds for β_{k+1}.

For $k + 1$, we have,

$$\beta_{k+1} = k + 1 - \sum_{i=1}^{k} \binom{k+1}{i} \beta_i.$$

Using $\binom{k+1}{m} = \binom{k}{m} + \binom{k}{m-1}$, $\forall m, k \in \mathbb{Z}_+$, reduces the problem to the form,

$$\beta_{k+1} = k + 1 - \sum_{i=1}^{k-1} \binom{k}{i} \beta_i - \beta_k - \sum_{i=1}^{k} \binom{k}{i-1} \beta_i.$$

Replacing the term, $\sum_{i=1}^{k-1} \binom{k}{i} \beta_i$ with $k - \beta_k$ and simplifying gives,

$$\beta_{k+1} = 1 - \sum_{i=1}^{k} \binom{k}{i-1} \beta_i.$$

For example, when $k = 2$ we get, $\beta_3 = 1 - \binom{2}{0}\beta_1 + \binom{2}{1}\beta_2$. Substituting $\beta_0 = 1$ and $\beta_2 = 0$ gives the required results.

Since by hypothesis, $\beta_i = 0$, for $i = 0, 2, 3 \ldots k$, and $\beta_1 = 1$, the result follows that $\beta_{k+1} = 1 - \beta_1 = 0$. Therefore, the result follows by induction. □

If $\lambda_L \to \infty$, then $\phi(\lambda_L) = E(\exp(-\lambda_L T_I) \to 0$, for all positive random variables T_I and the expression $\sum_{i=1}^{k-1} \binom{k}{i} \beta_i \phi(\lambda_L i)^{k-i}$, gives $\beta_k = k$ when $\lambda_L \to \infty$.

In figures 2 (a) and (b), we have plotted the β_k, as a function of λ_L while holding other parameters as $n = 6$, Gamma(a)=2, Gamma(b)=2.05 and $c = 1$ number of initial infective, for two extreme values of λ_L, that is when $\lambda_L \to \infty$ and when $\lambda_L \to 0$. We have adopted the Gamma infectious period distribution to enables us compare our results with those of [1] who also employed the Gamma(2, 2.05) infectious period distribution.

The behaviour of β_k is found to be consistent with our theoretical studies. When λ_L becomes very large, β_k becomes asymptotic to k, while as λ_L approaches 0, so also is β_k. This can be seen from figures 2.

In figure 2, we plotted the beta function as a function of λ_L, using Gamma(a, b) infectious period distribution with parameters Gamma$(a) = 2$, Gamma$(b) = 2.05$. We see that with increasing λ_L, the function β_k also increases and tends to $k = 1, \ldots, 5$, where $\beta_0 = 0$, while as λ_L tends to zero, β_k also tends to zero except β_1 which assumes the value 1.

3.3. The mean final size of the single household epidemic for small λ_L.

Using the properties of β_k and since $\phi(\lambda_L) \to 1$, if $\lambda_L \to 0$, the expression for the mean final size reduces to

$$\mu_{n,a} = n + a - \sum_{k=0}^{n} \binom{n}{k} \beta_k,$$

where $n + a$ is the household size, n and a are the number of initial susceptibles and infectives.

Since $\beta_k \to 0 \ \forall k \in \mathbb{Z}_+ - \{1\}$ with $\beta_1 = 1$, when $\lambda_L \to 0$, the expression for the mean final size reduces to,

$$\mu_{n,a} = n + a - \binom{n}{0} \beta_0 + \binom{n}{1} \beta_1.$$

Putting the values of $\beta_0 = 0$ and $\beta_1 = 1$ into the expression yields the value of the mean final size of a single household epidemic, when $\lambda_L \to 0$,

$$\mu_{n,a} = n + a - n = a.$$

This means that if there are no local contacts between susceptible and infective individuals in the household, there will be no new infections and the ultimate number of infected individuals at the end of the epidemic will be the initial number of infectives.

3.4. The mean final size of the single household epidemic, for large local infection rates.

If $\lambda_L \to \infty$, then $\phi(\lambda_L) = E(\exp(-\lambda_L T_I)) \to 0$, since T_I is a non-negative random variable and since β_k assumes the values $k \in \mathbb{Z}_+$, we can write the mean final size equation as,

$$\mu_{n,a} = n + a - \left(\binom{n}{0} \beta_0 \phi(\lambda_L.0)^{n+a-0} + \binom{n}{1} \beta_1 \phi(\lambda_L.1)^{n+a-1} + \cdots + \binom{n}{k} \beta_k \phi(\lambda_L.k)^{n+k-1} \right).$$

We know that if $\lambda_L \to \infty$, then $\phi(\lambda_L) \to 0$. The question then is, can $n + a - k$ be zero, since if $n + a - k$ is zero then the expression $\phi(\lambda_L.k)^{n+a-k}$ reduces to 1. Since k is only defined for $k = 0, 1, 2, \ldots, n$ and a is not zero, if a is zero then there will be no infection in the household and so no susceptible individuals will be subjected to any infection pressure and so $k < n + a$, $\forall a \in \mathbb{Z}_+ - \{0\}$,

However, if $k = 0$ then $\beta_k \phi(\lambda_L.k)^{a+n-k}$ reduces to zero, since $\beta_0 = 0$.

If $a \neq 0$, $k \neq 0$, then $n + a \ > k$.

Under this assumption, $\beta_k \phi(\lambda_L.k)^{a+n-k} \to 0$ and the summation terms on the right hand of mean final size will collapse to zero with the mean final size given by the remaining term as,

$$\mu_{n,a} = n + a.$$

This means that everybody will be infected at the end of the epidemic outbreak, which is possible for highly infectious diseases with large local contact rate. The role of these parameters on household disease transmission is crucial and any effective intervention, and control strategies must take this into consideration.

3.5. The threshold parameter for small and large local infection rates.

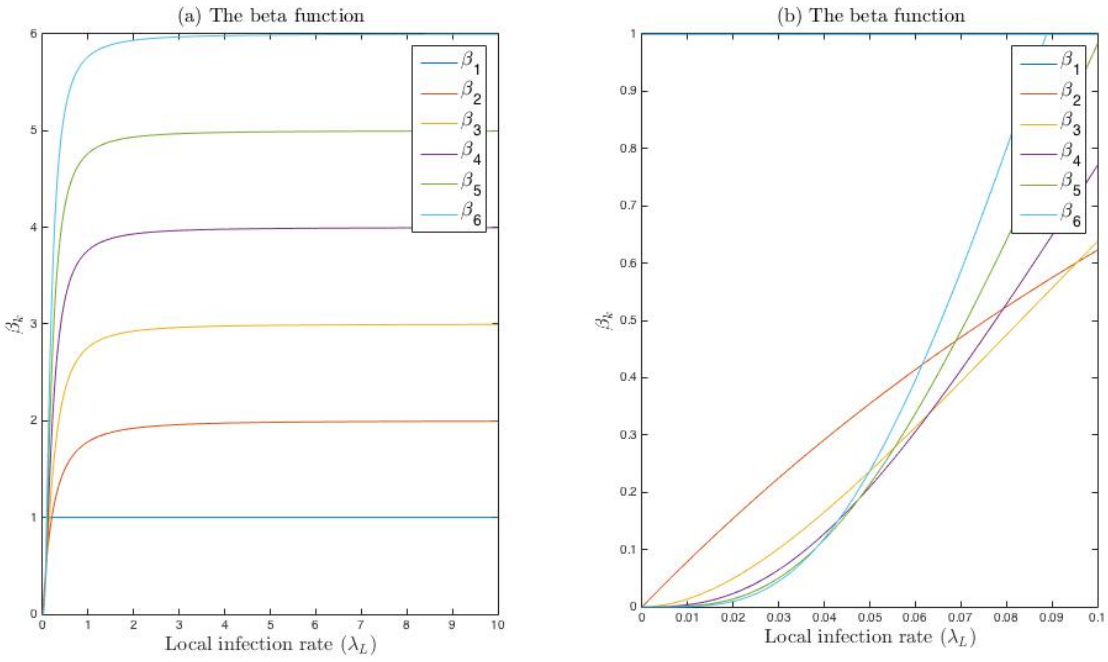

Figure 2: The beta function with increasing λ_L.

The threshold parameter is a function of both the local and global infection rates. If the global infection rate, $\lambda_G \to 0$, then the threshold parameter will be zero, on the contrary if $\lambda_L \to 0$, then β_k will all be zero except $\beta_1 = 1$ in accordance with the properties of β_k and the resulting mean final size $\mu_{n-1,1}$ of the household with $n-1$ initial susceptibles and 1 initial infective will be the initial infective, which under this definition is $\mu_{n-1,1} = 1$ with the threshold parameter given by

$$R_* = \lambda_G E(T_I) \sum_{n=1}^{\infty} \tilde{\alpha}_n,$$

Since $\tilde{\alpha}_n$ are probabilities, their summation will be 1, reducing the threshold parameter to

$$\begin{aligned} R_* &= \lambda_G E(T_I), \\ &= R_0. \end{aligned}$$

The household threshold parameter R_* is expressed in terms of R_0 in [4, 5, 6] as,

$$R_* = R_0 \mu,$$

where $R_0 = \lambda_G E(T_I)$ and $\mu = \sum_{n=1}^{\infty} \tilde{\alpha}_n \mu_n$ is the mean amplification factor owing to internal spread within the household. Where R_0 is the basic reproductive ratio for homogeneous mixing population, in which everyone is assumed to have similar characteristics without consideration for heterogeneity in infectivity and susceptibility. It is a threshold parameter for a population in which the household size is one. It can loosely be defined as the average number of infectives generated by a single infected individual in a completely susceptible population throughout its infectious period.

The behaviour of the threshold parameter for varying local infection is studied for some global infection rates, $\lambda_G = 0.01, 0.02, 0.03, 0.04$, and $\lambda_G = 0.1, 0.2, 0.3, 0.4$ respectively, Gamma(a, b) infectious period distribution with parameters, $a = 2, b = 2.05$ and assuming single initial infective, $c = 1$, in the household. We found in each of the cases that large global infection rate leads to corresponding large threshold parameter. Thus the threshold parameter is linearly influenced by the level of the global contact rate.

In figure 3, we have plotted the threshold parameter for varying local infection rate defined in the region $\{\lambda_L : 0 \leq \lambda_L \leq 2\}$, with stepsize of 0.05, for the following global infection rates $\lambda_G = 0.01, 0.02, 0.03, 0.04$, and $\lambda_G = 0.1, 0.2, 0.3, 0.4$ respectively and Gamma$(2, 2.05)$ infectious period distribution, and one initial infective, $c = 1$.

4. Proportion of the initial susceptibles that are ultimately infected.

The proportion of the initial susceptible individuals that are ultimately infected by the epidemic, denoted by z, is given in [6] as

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} \mu_{n-k,k}. \tag{2}$$

Equation (2) is the weighted average of the number of infectives in a single household epidemic with Binomial distributed number of infectives k, and the remaining $n-k$ susceptibles avoid infection from outside the household of size n.

In equation (2), $\tilde{\alpha}_n$ is the probability that a randomly selected individual resides in a household of size n, π is the probability that a given individual avoids global infection, which is approximately given in [5, 6] as,

$$\pi = \exp(-\frac{\lambda_G}{N} z N E(T_I)) = \exp(-\lambda_G z E(T_I)). \tag{3}$$

Where $NzE(T_I)$ is the total person units of infection present throughout the epidemic, N is the total number of individuals in the household and z is the proportion of the initial susceptibles ultimately infected.

Suppose global epidemic has occurred with the proportion of individuals ultimately infected, $z \in [0, 1]$, then equations (2) together with (3) gives an implicit equation for z. Here $z = 0$ is always a solution and the only solution if $R_* \leq 1$. A second solution in $0 < z < 1$ exists only if $R_* > 1$. This is better understood by expressing equation (3) in the form $y = z = g(z)$ where $y = z$, $y = g(z)$.

Here $g(z)$ is the right hand side of equation (3) and the unique solution of the equation is found at the point of intersection of $y = z$ and $y = g(z)$ nearest to the origin for which $R_* > 1$. Now let the generating function of the offspring random variable R be defined as $E(z^R) = g(z)$ and P_k be its distribution. Then $g(z) = \sum_{k=0}^{\infty} P_k z^k$ with $g'(1) = \sum_{k=1}^{\infty} k P_k$ which is equal to R_*.

4.1. Proportion of the initial susceptibles that are ultimately infected at the lower and upper boundaries of the local infection rate.

If the local contact rate $\lambda_L \to 0$, then the mean final size of a household with k initial infectives and $n-k$ initial susceptibles is, $\mu_{n-k,k}(0) = k$. We can express z as,

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} k. \tag{4}$$

Since,

$$E(K) = \sum_{k=0}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} k = n(1-\pi),$$

where $\binom{n}{k}(1-\pi)^k \pi^{n-k}$ is the probability that k susceptibles individuals are infected with probability $(1-\pi)^k$, while the remaining $n-k$ escape infection with probability π^{n-k}.

The number of infectives k, in the household is distributed as a binomial random variable, with parameters, n and $(1-\pi)$.

Here $E(K)$ is the mean number of infected susceptibles in the household. Substitution of the mean number of susceptible individuals infected, $E(K) = n(1-\pi)$ into the expression for z gives,

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n (1-\pi) = (1-\pi).$$

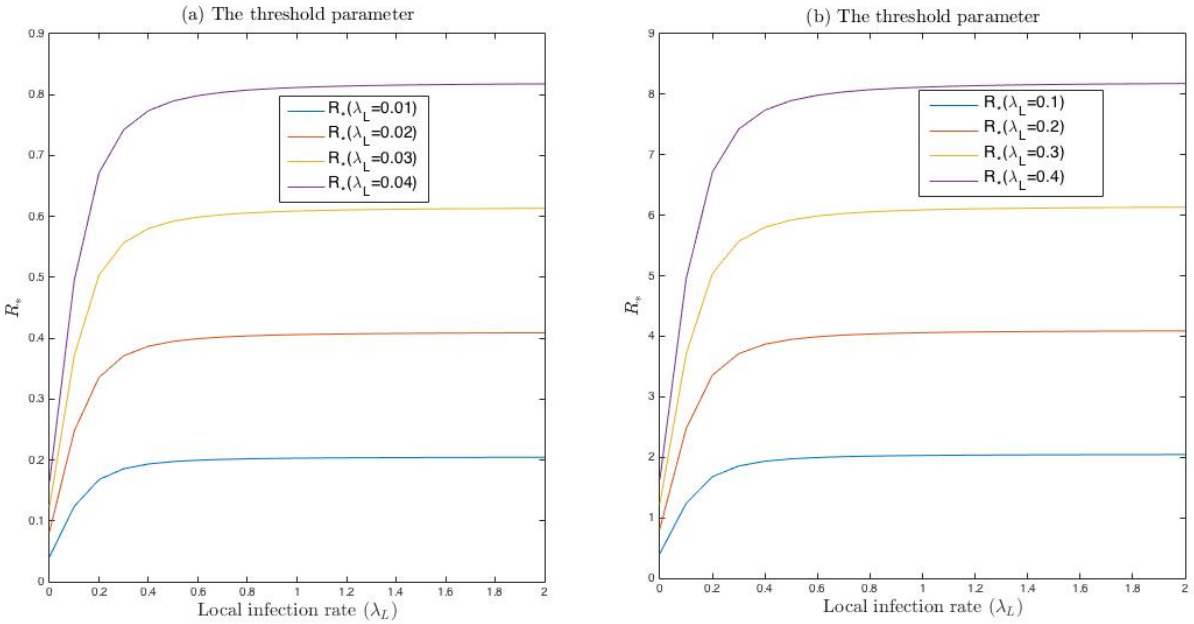

Figure 3: The threshold parameter with varying local infection rate.

This can further be simplified as,

$$z = 1 - \pi = 1 - \exp(-\lambda_G z E(T_I)). \tag{5}$$

This is the governing equation of z for the single population S-I-R deterministic epidemic model.

If $\lambda_L \to \infty$, then the mean final size in equation, $\mu_{n-k,k}$ for $k > 0$ reduces to n and the expression for z becomes,

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} n. \tag{6}$$

Since,

$$\sum_{k=0}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} = 1,$$

we will have,

$$\sum_{k=0}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} = \pi^n + \sum_{k=1}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k},$$

where $p(K=0) = \pi^n$ is the probability that every susceptible in a household of size n avoids global inflection. We can write

$$\sum_{k=1}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} = 1 - \pi^n.$$

We can then express z as,

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^{n} \binom{n}{k} (1-\pi)^k \pi^{n-k} n = \sum_{n=1}^{\infty} \tilde{\alpha}_n (1 - \pi^n),$$

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n (1 - \exp(-n\lambda_G z E(T_I))),$$

where $\pi^n = \exp(-n\lambda_G z E(T_I))$. Further simplification of z gives,

$$z = 1 - \sum_{n=1}^{\infty} \tilde{\alpha}_n \exp(-n\lambda_G z E(T_I)). \tag{7}$$

5. Estimation in the presence of model misspecification.

Here, the estimation methods uses the maximum likelihood techniques as described in [1] and [22], with starting values using [12] method described in [22]. These are implemented in subsections 5.1 and 5.2 respectively.

5.1. Simulating epidemic data with exp(4.1) and estimating model parameters with Gamma(2, 4.1/2) infectious period distributions.

We simulate two dimensional model epidemic data with exp(4.1) infectious period distribution and estimated the model parameters with the Gamma(2, 4.1/2) infectious period distribution. Plots of the estimates and tables of mean, standard deviation and root mean square errors are presented.

From figures 4 (a)-(d), we see that the estimates are biased and imprecise.

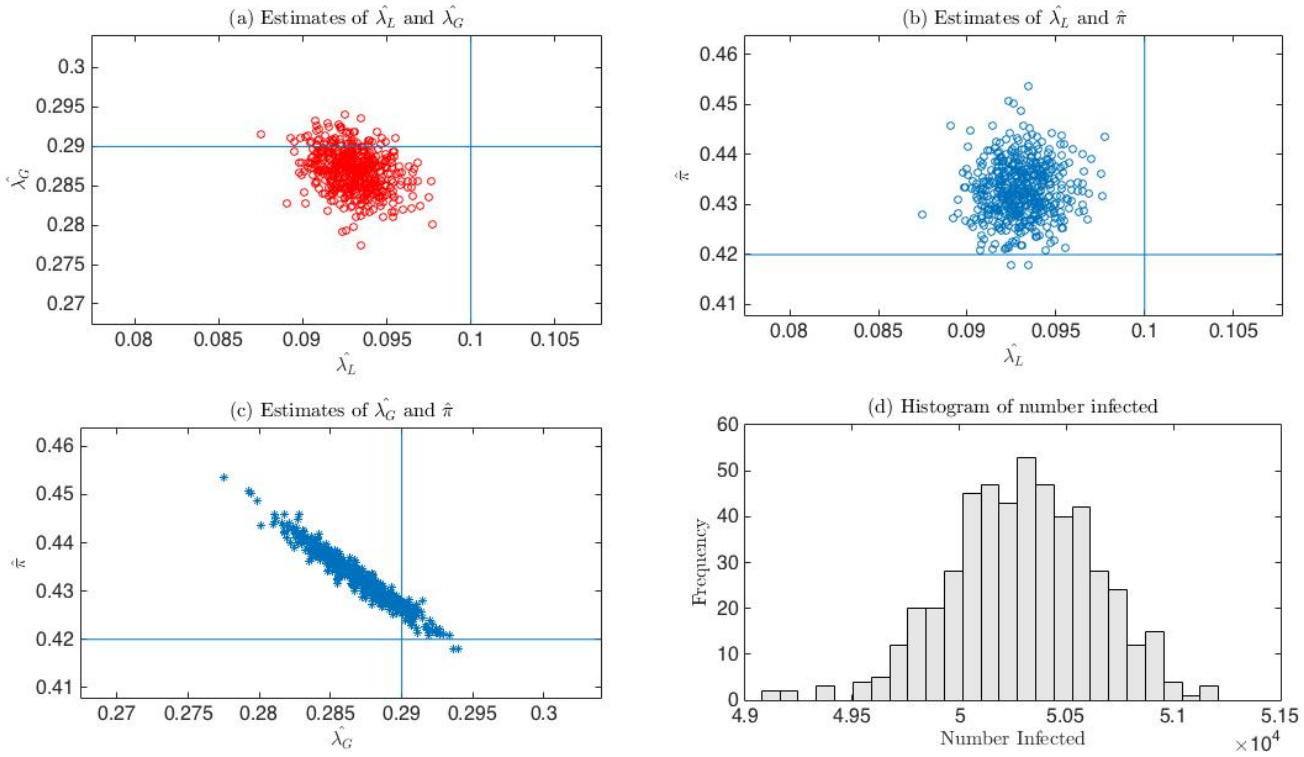

Figure 4: Plots of the estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(1.4) infectious period distribution.

	Gamma$(2, 4.1/2)$ infectious period distribution				
Par. Mean SD, RMSE	λ_L	λ_G	π	z	R_*
Theoretical parameter	0.1	0.29	0.4199	0.7298	2.2166
Mean	0.092993	0.2869	0.43285	0.7119	2.1339
Standard deviation	0.0015132	0.0026017	0.005495	0.0048041	0.019735
Root mean square error	0.0071679	0.0040445	0.014025	0.018455	0.084961

Table 1: Table of mean, standard deviation and root mean square error of the estimates when the epidemic data is simulated with exp(4.1) and estimated with Gamma$(2, 4.1/2)$ infectious period distributions.

	exp(4.1) infectious period distribution				
Par. Mean SD, RMSE.	λ_L	λ_G	π	z	R_*
Theoretical parameter.	0.1	0.29	0.4291	0.7117	2.1106
Mean	0.10761	0.29351	0.41595	0.72898	2.1878
Standard deviation	0.0019244	0.0023979	0.0047063	0.0039479	0.017285
Root mean square error	0.0078474	0.0042457	0.013884	0.017705	0.079101

Table 2: Table of mean, standard deviation and root mean square error of the estimates when the epidemic data is simulated with Gamma$(2, 4.1/2)$ and estimated with exp(4.1) infectious period distributions.

5.2. Simulating epidemic data with Gamma$(2, 4.1/2)$ and estimating model parameters with exp(4.1) infectious period distributions.

Here, we estimate the model parameters with exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1)$ infectious period distribution.

Plots of the parameter estimates, table of mean, standard and root mean square of the estimates are presented as follows.

In figures 5 (a)-(d), the estimates are biased and imprecise.

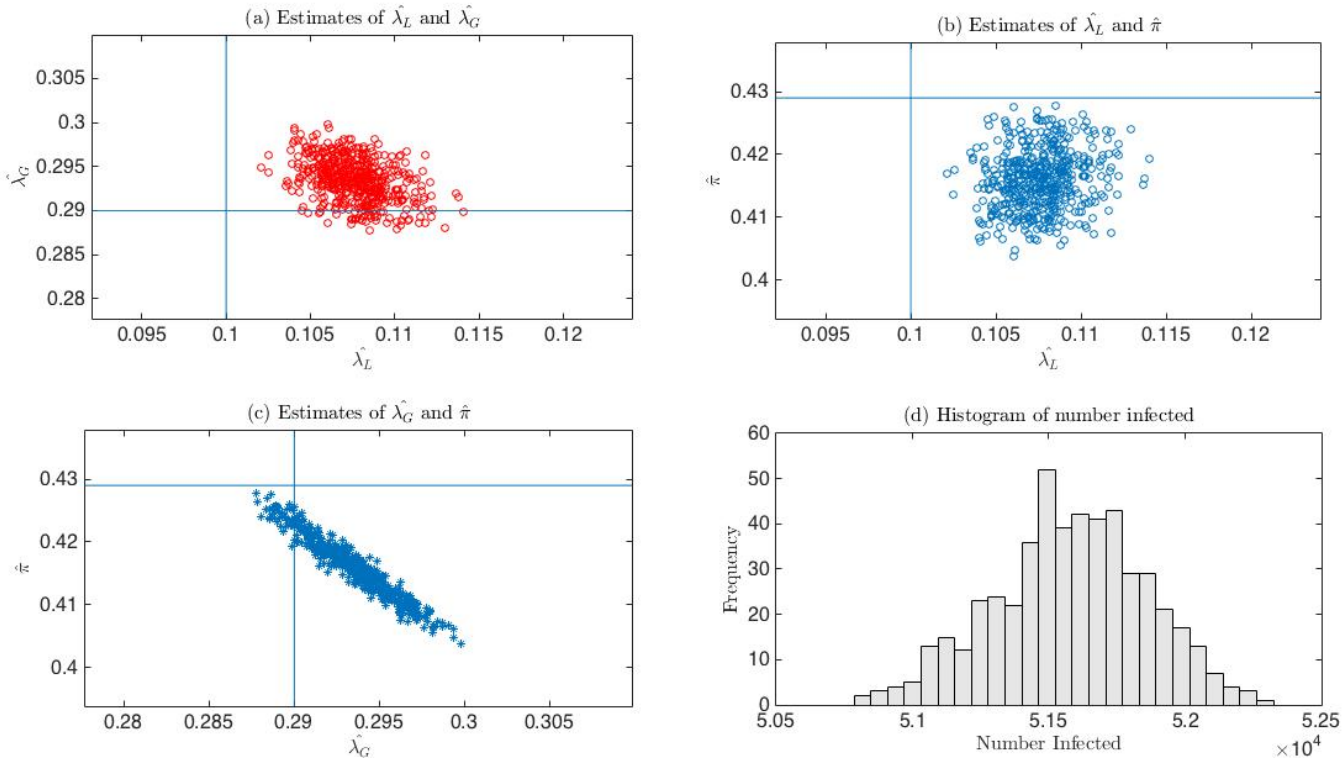

Figure 5: Plots of the estimates with exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution.

6. Results and Discussion.

From figures 4 (a)-(d), we see that the estimates are biased and imprecise if wrong infectious period distribution is chosen for model estimation. These behavours is not inconsistent with the results of our studies in [22] with large choice of the minimum epidemic size 1000 say and suitable theoretical parameters in which global epidemic occurred. We find that parameter estimates are scattered around their true values as expected compared to the behaviours displayed in figures 4 and 5, (a)-(d) respectively, which are contrary to the expected results.

Tables 1 and 2 of mean, variance, mean square error and root mean square error of the estimates are indicative of the quality of the estimates given this scenario.

7. Conclusion

This work has shown that, there is the need for consistency in the choice of the infectious period distribution, as this has considerable influence on the model estimation and fitting to the final size data. Wrong choice of the distribution leads to a model which does not well to the final size data.

Acknowledgments

The Author will like to acknowledge Dr. Owen D. Lyne, and Professor Martin Ridout for their valuable contributions and guildance during my research training at the University of Kent, Canterbury, UK.

References

[1] C. ADDY, I. M. LONGINI JR, AND M. HABER, A Generalised Stochastic Model for the Analysis of Infectious Disease Final Size Data. Biometrics, Vol. 47, No. 3, (1991), pp. 961-974.

[2] H. ANDERSSON AND T. BRITTON, Lecture Notes in Statistics: Stochastic Epidemic Models and Their Statistical Analysis. Springer, Verlag (2000).

[3] F. G. BALL, The Threshold Behaviour of Epidemic Models. Journal of Applied Probability, Vol. 20, No. 2, (1983), pp. 227-241.

[4] F. G. BALL, A Unified Approach to the Distribution of the total size and Total Area under the Trajectory of Infection in Epidemic Models. Advances in Applied Probability, Vol. 18, No. 2, (1986), pp. 289-310.

[5] F. G. BALL, D. MOLLISON AND G. SCALIA-TOMBA, Epdemics with Two Levels of Mixing. Annals of Applied Probability, Vol. 7, No. 1, (1997), pp. 46-89.

[6] F. G. BALL AND O. D. LYNE, Epidemics Among A Population of Households. Mathematical Approaches for the Emerging and Reemerging Infectious Disease: Models, Methods and Theory, (The IMA Volumes in Mathematics and its Applications), Springer, Editor: Castillo-Chavez, Vol. 126, (2000), pp. 115-125.

[7] F. G. BALL AND O. D LYNE, Stochastic Multitype SIR Epidemic Among A Population Partitioned into Households. Advances in Applied Probability, Vol. 33, No. 1 (Mar. 2001), pp. 99-123

[8] F. G. BALL AND P. NEAL, A general model for the stochastic SIR epidemic with two levels of mixing. Journal of Math. Biosciences, Vol. 180, (2002), pp. 73-102.

[9] F. BALL AND P. NEAL, Network Epidemic Models With Two Levels of Mixing. Mathematical Biosciences, Vol. 212, (2008), pp. 69-87.

[10] , R. BARTOSZYNSKI, Branching Processes and the Theory of Epidemic. Fifth Berkeley symposium on Mathematical Statistics and Probability, University of California Press, Vol. 4, (1967), pp.259-269.

[11] , R. BARTOSZYNSKI, On a Certain Model of An Epidemic. Journal of Applicatione Mathematicae, Vol. 13, (1972), pp. 139-151.

[12] N. G. BECKER, A stochastic Model for Interacting Population. Journal of Applied Probability, Vol. 7, No. 3, (1970), pp. 544-564.

[13] N. G. BECKER, Analysis of Infectious Disease Data: Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, (1989).

[14] F. BRAUER, P. VAN DEN DRIESSCHE AND J. WU (EDS.) Mathematical Epidemiology, Mathematical Biosciences Subseries, Springer-Verlag, Berlin (2008).

[15] D.J.DALEY AND J.GANI, Epidemic Modelling: An Introduction. Cambridge University Press, New York, (1999).

[16] O. DIEKMANN, H. HEESTERBEEK AND T. BRITTON, Mathematical Tools for Understanding Infectious Diseases Dynamics: Princeton Series in Theoretical and Computational Biology. Princeton University Press, New Jersey, (2013).

[17] C. LEFEVRE AND P. PICARD, A Non-Standard Family of Polynomial and The Final Size Distribution of Reed-Frost Epidemic Processes. Advances in Applied Probability, Vol.22, No.1, (1990), pp. 25-48.

[18] I. M. LONGINI, JR AND J.S. KOOPMAN, Household and Community Transmission Parameters from Final Distribution of Infections in Households. Biometrics, Vol. 38, No. 1, (1982), pp. 115-126.

[19] I. M. LONGINI, JR, J. S. KOOPMAN, A. S. MONTO, AND J. P. FOX, Estimating Household and Community Transmission Parameters for Influenza. American Journal of Epidemiology, Vol. 115, No. 5, (1982), pp. 736-750.

[20] P. NEAL, Efficient Likelihood-free Bayesian Computation for Household Epidemics . Journal of Statistics and computing, Vol. 22, No.6, (2012), pp. 1239-1256.

[21] P. NEAL, A Household SIR Epidemic Model Incorporating Time of Day Effects. Journal of Applied Probability, Vol. 53, (2016), pp. 489-501.

[22] A. M. UMAR, Effects on mimimum epidemic and population sizes on a global epidemic in simulation of final size epidemic data. Journal of progressive research in Mathematics, Vol. 16, (2020), pp. 3093-3108.