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

Inference of the Stochastic SIR Household Epidemic Model with Misspecification and Misclassification

Umar M.A

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

Received: June 21, 2021; Accepted: August 17, 2021; Published: September 8, 2021

Abstract

Model estimate and their functions are affected by wrong choice of the infectious period distribution, T_I when the actual one is unknown. This is a misspecification problem which is often accompanied with biased and imprecise estimates as discussed in [16], which may be taking for misclassification of the final size epidemic data . This work, examined these problems using simulations by assuming constant infectious period, $T_I \equiv 4.1$ and then estimated with T_I infectious period distribution, assumed as $\Gamma(2, 2.05)$ for the household epidemic and vice versa under. This is extended to cases when the final size data is misclassified, as studied in [16]. The maximum likelihood estimates and the model fitness to the final size data are examined and compared under these conditions. We found that, in the two cases, the estimates are biased and imprecise. Thus, model that fit poorly to the final size data is often obtained. However, the choice of appropriate model that fit better to the final size data, given these scenarios are suggested.

Keywords:

Final size epidemic, infectious period distribution, misspecification, misclassification probabilities.

1. Introduction

This work examined two scenarios namely, model estimates when the final size data is misclassified and the infectious period distribution is misspecified in parameter estimation. We investigate these behaviours and provide suggestions on possible choice of appropriate model which fit the final size data better. We do this using simulations with large population size and appropriate choice of the theoretical parameters which allows global infection. The computation methods is can be found in [1], [10], [13], [12]and [14] with starting values proposed by [10] also discussed in [15]. This is done by constructing the approximate likelihood function, which uses the final size probabilities, found in [1], [7], [8, 2] and [15]. The model assume the

How to cite this paper: A. M. Umar (2021) Inference of the Stochastic SIR Household Epidemic Model with Misspecification and Misclassification. Journal of Progressive Research in Mathematics, 18(3), 60-81. Retrieved from http://scitecresearch.com/journals/index.php/jprm/article/view/2075

household structure in [7], [8, 2]. Other properties of this model are examined in [3], [4], [5], [6] and [15].

1.1. Misspecification of the Infectious period distribution

Misspecification often occur when a different infectious period distribution from that used in simulation is employed in estimation. This may wrongly be assumed as a misclassification problem, as model estimates are often biased and imprecise as observed in [16], hence model that does not fit significantly well to the final size data is obtained. There is therefore, the need to study the effects of this scenario on the model estimates and also compare them with situation when the final size data is misclassified. We examined these problems, using simulations as in sections 2 and 3 respectively hence can be compared with the behaviours in [16]. Plots of the estimates, their mean, standard deviation and root mean square error are presented to give further insights into the model fitness to the final size epidemic data, under these scenarios.

1.2. Misclassified final size epidemic data.

Measurement error occurs when the real variable is unavailable and replaced by its surrogate often referred to as naive [11]. For example, in a regression analysis with explanatory variable X and response Y, either of the variables can be subject to mismeasurement. On categorical data, mismeasurement occurs when the actual and recorded categories for subjects differ. Here, the surrogate variable cannot be expressed as sum of the true variable plus a noise variable, rather they are expressed in terms of classification probabilities called misclassification probabilities.

Let x and y be the observed false and true positives in a household of size n, then the probability of observing $x+y=i$ positives, given that the true number of positives is j can be written as,

$$P_{i,j}(n) = P(x+y=i \,|\, \text{True infect} = j, \text{household size} = n). \tag{1}$$

The probability of observing $i \in \mathbb{Z}_+ \leq n$ infectives in a household of size n, has been shown in [17] to be,

$$P_{r,j}(n) = \sum_{k=0}^{r} \binom{j}{r-k} \binom{n-j}{k} \varepsilon_{FN}^{j-r+k} (1-\varepsilon_{FN})^{r-k} \varepsilon_{FP}^{k} (1-\varepsilon_{FP})^{n-j-k} \tag{2}$$

where, ε_{FP} and ε_{FN}, are the false positive and negative probabilities.

Alternatively, written as,

$$P_{r,j}(n) = \sum_{k=0}^{r} \binom{j}{k} \binom{n-j}{r-k} \varepsilon_{FN}^{j-k} (1-\varepsilon_{FN})^{k} \varepsilon_{FP}^{r-k} (1-\varepsilon_{FP})^{n-j-r+k}. \tag{3}$$

Where

$$\sum_{i=0}^{n} P_{i,j}(n) = 1, \, \forall j \in \mathbb{Z}_+ \leq n.$$

If $\varepsilon_{FN} = \varepsilon_{FP} = \varepsilon$, then $P_{i,j}(n)$, $i,j = 0,1,\ldots,$ (2) simplifies to

$$P_{i,j}(n) = \sum_{k=0}^{i} \binom{j}{i-k} \binom{n-j}{k} \varepsilon^{j-i+2k} (1-\varepsilon)^{n-j+i-2k}, \quad i,j = 0,1,\ldots,n. \tag{4}$$

Alternatively,

$$P_{i,j}(n) = \sum_{k=0}^{i} \binom{j}{k} \binom{n-j}{i-k} \varepsilon^{j+i-2k} (1-\varepsilon)^{n-j-i+2k}, \quad i,j = 0,1,\ldots,n. \tag{5}$$

satisfying,

$$\sum_{i=0}^{n} P_{i,j}(n) = 1, \, \forall j \in \{0,1,\ldots,n\}.$$

The approximate likelihood function has the form,

$$L(\lambda_L, \pi, \varepsilon_{FP}, \varepsilon_{FN}) \propto \prod_{n=1}^{max} \prod_{i=0}^{n} q_{n,i}(\lambda_L, \pi, \varepsilon_{FP}, \varepsilon_{FN})^{x_{n,i}}. \tag{6}$$

where max is the maximum household size.

Expressed using log likelihood function as,

$$\ell(\lambda_L, \pi, \varepsilon_{FP}, \varepsilon_{FN}) = \sum_{n=1}^{max} \sum_{i=0}^{n} \left(x_{n,i} \log_e \left(\sum_{j=0}^{n} P_{i,j}(n) P_j(n) \right) \right), \; i, \; j = 0, 1, \cdots, n. \tag{7}$$

Where $\log(L(\lambda_L, \pi, \varepsilon_{FP}, \varepsilon_{FN})) = \ell(\lambda_L, \pi, \varepsilon_{FP}, \varepsilon_{FN})$

For the three dimensional model, ε_{FP}=ε_{FN} in the approximate function.

2. Inference of Misspecification with Misclassified Final Size epidemic Data.

We studied the problems using large population size and theoretical parameters which allows global infection in our simulations as in [15], by considering the pair of theoretical parameters, $(\lambda_L, \lambda_G) = (0.1, 0.29)$ and a range of misclassification probabilities, $\varepsilon = 0.01, 0.02, 0.2$.

Plots of the estimates with the two scenarios, in which the epidemic data is estimated with a different infectious period from that used in simulating the data are presented as follow.

2.1. When the epidemic data is simulated with exp(4.1) and estimated with Gamma(2, 4.1/2) infectious period distributions.

Here we simulate the epidemic data with exp(4.1) infectious period distribution and estimate the model parameters with Gamma(2, 4.1/2) infectious period distribution as follow.

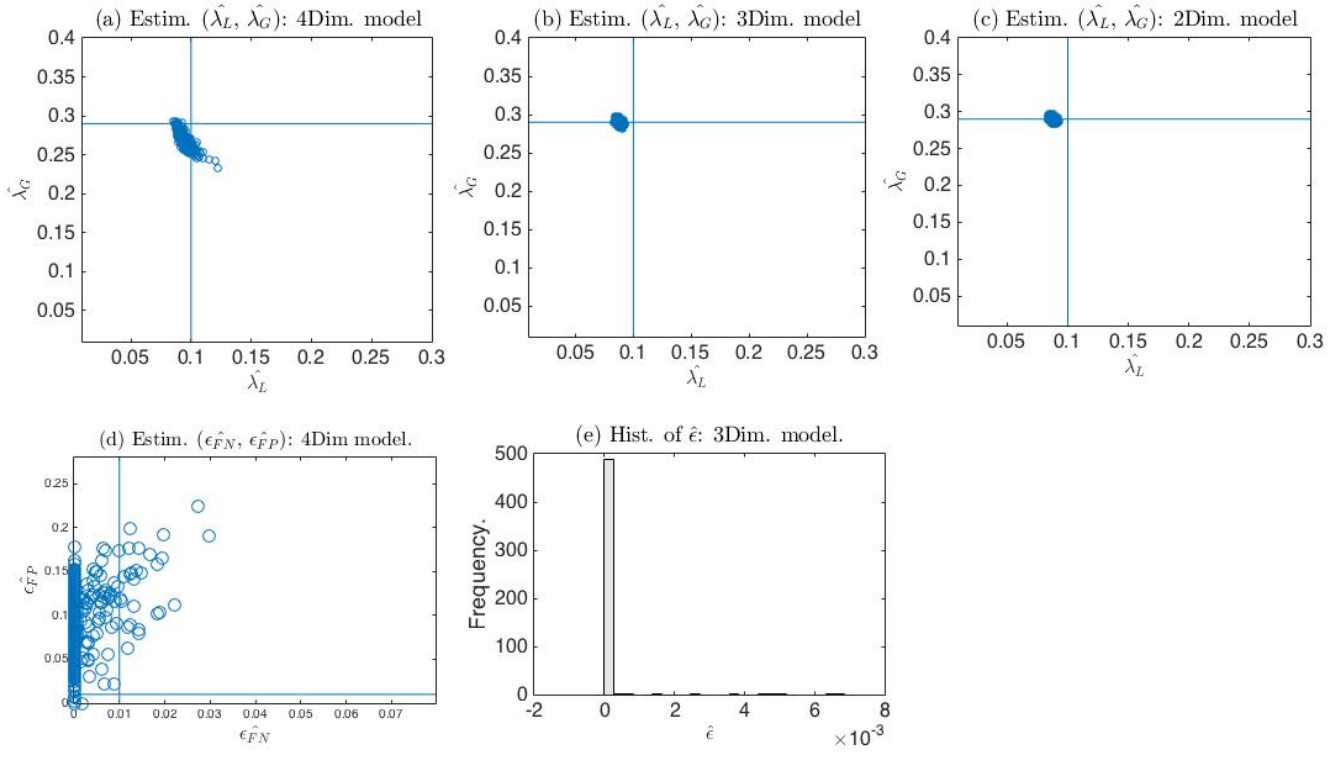

Figure 1: Plots of the estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution and $\varepsilon = 0.01$.

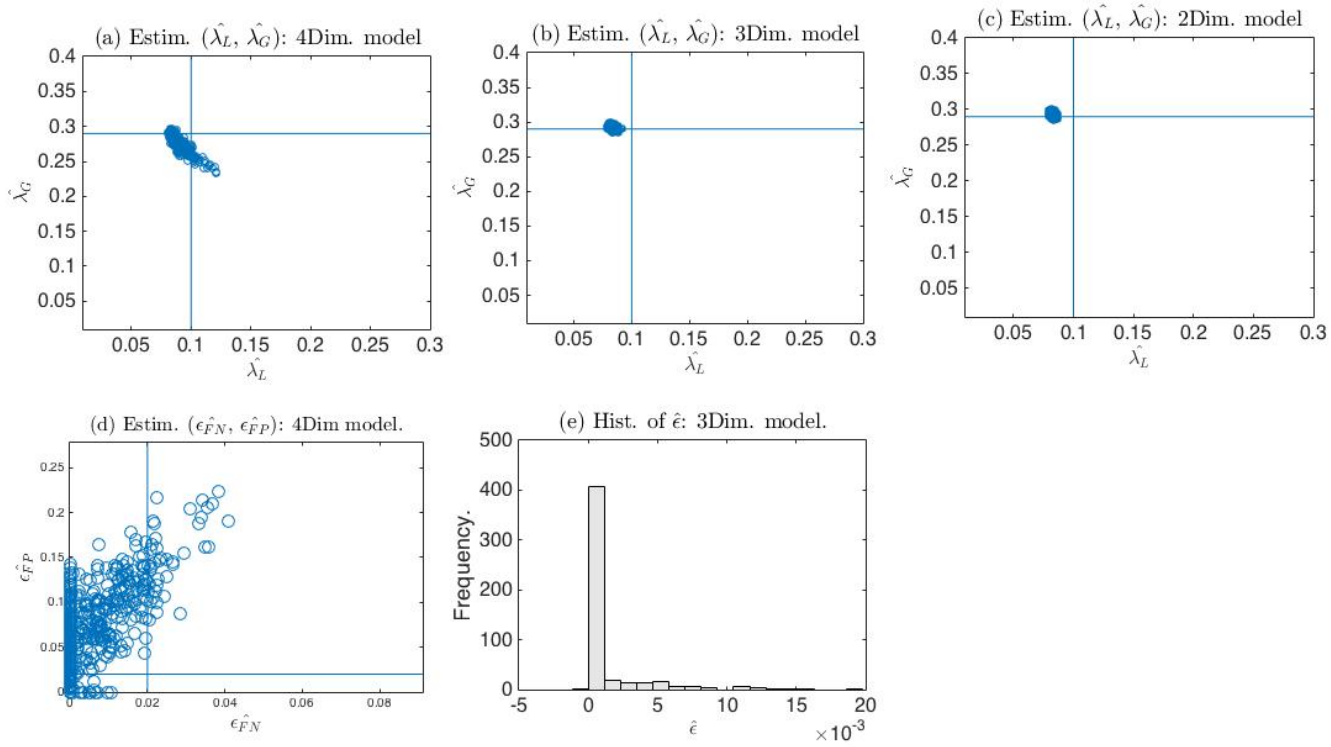

Figure 2: Plots of the estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution and $\varepsilon = 0.02$.

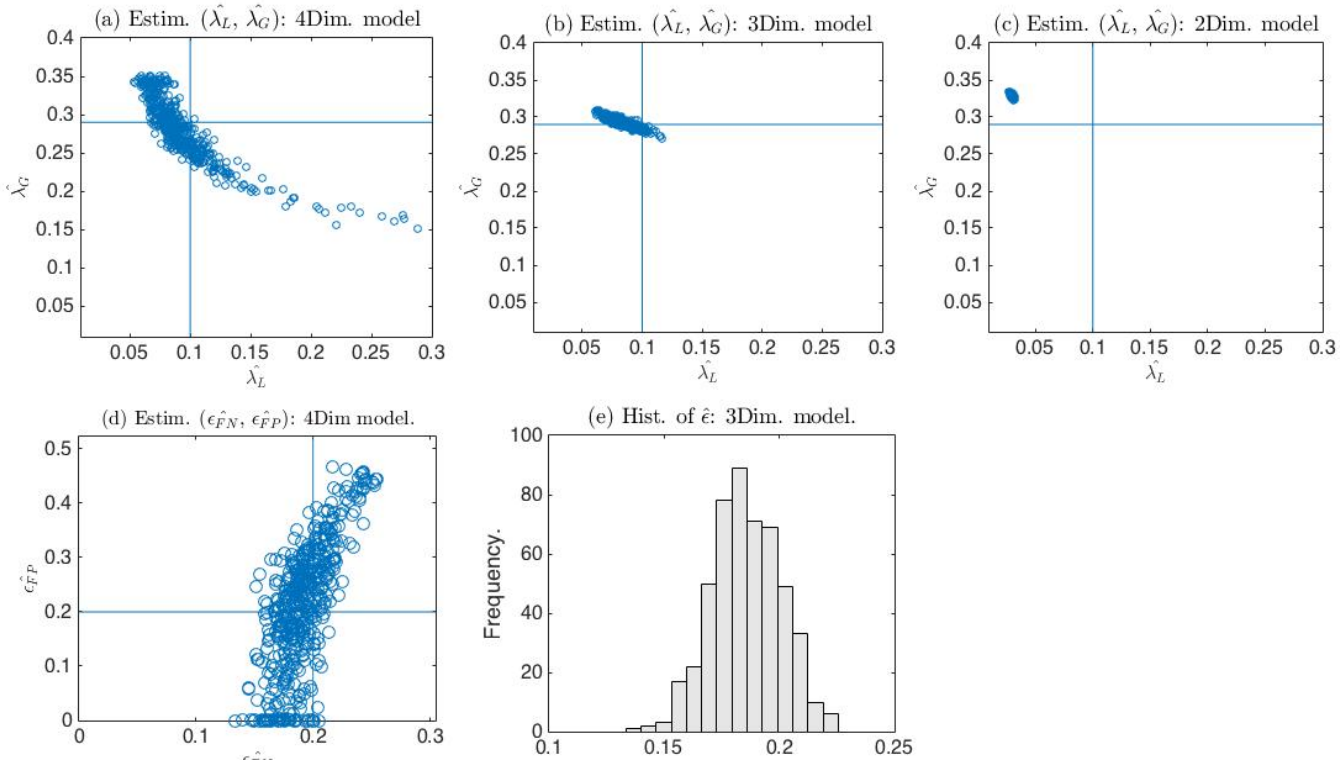

Figure 3: Plots of the estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution and $\varepsilon = 0.2$.

	Model									
	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	Theor.
ε	0.01	0.01	0.01	0.02	0.02	0.02	0.2	0.2	0.2	Par.
$\hat{\lambda_L}$	0.087924	0.087965	0.093676	0.083014	0.083485	0.089956	0.029568	0.082831	0.10262	0.1
$\hat{\lambda_G}$	0.28989	0.28987	0.26986	0.29261	0.29234	0.27534	0.32819	0.29288	0.27823	0.29
$\hat{\pi}$	0.43114	0.43115	0.4721	0.43014	0.43025	0.46311	0.43059	0.4303	0.46004	0.4199
$\hat{z}$	0.70788	0.70792	0.67864	0.70323	0.70367	0.6823	0.62622	0.70248	0.68729	0.7298
$\varepsilon\hat{_{FN}}$	N/A	N/A	0.0014387	N/A	N/A	0.0062863	N/A	N/A	0.19136	N/A
$\varepsilon\hat{_{FP}}$	N/A	N/A	0.092036	N/A	N/A	0.077538	N/A	N/A	0.2094	N/A
$\hat{\varepsilon}$	N/A	8.33E-05	N/A	N/A	0.0010209	N/A	N/A	0.18576	N/A	N/A
$\hat{R_*}$	2.1115	2.1117	2.0115	2.0863	2.0887	2.0198	1.7083	2.0823	2.0392	2.2166

Table 1: Mean of the parameter estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the data is simulated with exp(4.1) infectious period distribution.

	Model								
	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim
ε	0.01	0.01	0.01	0.02	0.02	0.02	0.2	0.2	0.2
$\hat{\lambda_L}$	0.0014051	0.0014386	0.0040913	0.001363	0.0018231	0.0066354	0.00093879	0.0097516	0.065447
$\hat{\lambda_G}$	0.0025843	0.0026028	0.0086708	0.0024284	0.002515	0.010904	0.0020427	0.0061165	0.045795
$\hat{\pi}$	0.005432	0.0054391	0.017883	0.0049548	0.0049653	0.020602	0.0036992	0.0063326	0.084384
$\hat{z}$	0.0047356	0.0047228	0.012985	0.0042594	0.0044268	0.013337	0.0031339	0.010575	0.047823
$\varepsilon\hat{_{FN}}$	N/A	N/A	0.0039525	N/A	N/A	0.008371	N/A	N/A	0.02099
$\varepsilon\hat{_{FP}}$	N/A	N/A	0.036955	N/A	N/A	0.044611	N/A	N/A	0.1174
$\hat{\varepsilon}$	N/A	0.00062226	N/A	N/A	0.0026397	N/A	N/A	0.014962	N/A
$\hat{R_*}$	0.018951	0.018903	0.043454	0.016555	0.017794	0.043505	0.0079267	0.053799	0.16304

Table 2: Standard deviation of the parameter estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the data is simulated with exp(4.1) infectious period distribution.

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

We examined the properties of the estimates under these scenarios, presented their plots and tables of mean standard deviation and root mean square error.

3. Misspecification in the face of different Misclassification Probabilities.

Here, we studied the effect of misspecification on the estimate of the model parameters, when the epidemic data is misclassified with different misclassification probabilities, such that the infectious period distribution used in estimation is different from that used in simulating the epidemic data.

We examined this problem by simulating epidemic with theoretical parameters $(\lambda_L, \lambda_G) = (0.1, 0.29)$, Gamma$(2, 4.1/2)$ infectious period distribution for a range of $\{(\varepsilon_{FN}, \varepsilon_{FP}) : \varepsilon_{FN} \in [0, 1], \varepsilon_{FP} \in (0, 1)$ and then estimate the models with exp(4.1) infectious period distributions.

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

We simulate epidemic data with exp(4.1) infectious period distribution and estimate the model with Gamma$(2, 4.1/2)$ infectious period distribution. We present plots of the estimates and table of the mean,

	Model								
	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim
ε	0.01	0.01	0.01	0.02	0.02	0.02	0.2	0.2	0.2
$\hat{\lambda_L}$	0.012157	0.012121	0.0075297	0.01704	0.016615	0.012035	0.070438	0.01974	0.065434
$\hat{\lambda_G}$	0.002584	0.0026035	0.0075297	0.0035642	0.0034355	0.012035	0.038244	0.0067555	0.065434
$\hat{\pi}$	0.012492	0.012504	0.055177	0.011382	0.011484	0.047868	0.011316	0.012178	0.093372
$\hat{z}$	0.022415	0.022377	0.052767	0.026896	0.02649	0.049329	0.10362	0.029289	0.063947
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.0094462	N/A	N/A	0.016062	N/A	N/A	0.022678
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.009428	N/A	N/A	0.072779	N/A	N/A	0.11766
$\hat{\varepsilon}$	N/A	0.0099362	N/A	N/A	0.019161	N/A	N/A	0.020643	N/A
$\hat{R_*}$	0.10678	0.10658	0.20961	0.13136	0.12914	0.20151	0.50835	0.14461	0.24079

Table 3: Root mean square error of the parameter estimates with Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution.

	Model									
	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	Theor.
ε	0.01	0.01	0.01	0.02	0.02	0.02	0.2	0.2	0.2	Par.
$\hat{\lambda_L}$	0.10058	0.1274	0.12439	0.094141	0.12782	0.12418	0.030468	0.12778	0.14371	0.1
$\hat{\lambda_G}$	0.29647	0.28606	0.29078	0.29931	0.28595	0.29154	0.333	0.28622	0.28769	0.29
$\hat{\pi}$	0.41465	0.41957	0.41156	0.41355	0.41959	0.41022	0.41912	0.41963	0.42183	0.4291
$\hat{z}$	0.72426	0.74059	0.74488	0.71956	0.74085	0.74575	0.63694	0.74022	0.74092	0.7117
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.030553	N/A	N/A	0.040032	N/A	N/A	0.21275	N/A
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.010832	N/A	N/A	0.01722	N/A	N/A	0.18887	N/A
$\hat{\varepsilon}$	N/A	0.032266	N/A	N/A	0.042179	N/A	N/A	0.21254	N/A	N/A
$\hat{R_*}$	2.1617	2.2509	2.2698	2.1356	2.2522	2.2738	1.73	2.2486	2.2593	2.1106

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

	Model								
	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim
ε	0.01	0.01	0.01	0.02	0.02	0.02	0.2	0.2	0.2
$\hat{\lambda_L}$	0.0017218	0.0064309	0.0076548	0.0016489	0.0068501	0.0087742	0.00098638	0.016316	0.070344
$\hat{\lambda_G}$	0.0024134	0.0033416	0.0069713	0.0023772	0.0033419	0.0082493	0.0019164	0.0061721	0.04847
$\hat{\pi}$	0.0047786	0.0052086	0.012119	0.004571	0.0050424	0.014044	0.0033002	0.0065439	0.083907
$\hat{z}$	0.0040338	0.0051911	0.007914	0.0037568	0.0053171	0.008631	0.0027658	0.010044	0.044281
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.0064402	N/A	N/A	0.0069957	N/A	N/A	0.015604
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.027343	N/A	N/A	0.031997	N/A	N/A	0.12798
$\hat{\varepsilon}$	N/A	0.0063887	N/A	N/A	0.0065142	N/A	N/A	0.011181	N/A
$\hat{R_*}$	0.017093	0.025695	0.037354	0.01539	0.026481	0.040686	0.0074831	0.052057	0.19753

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

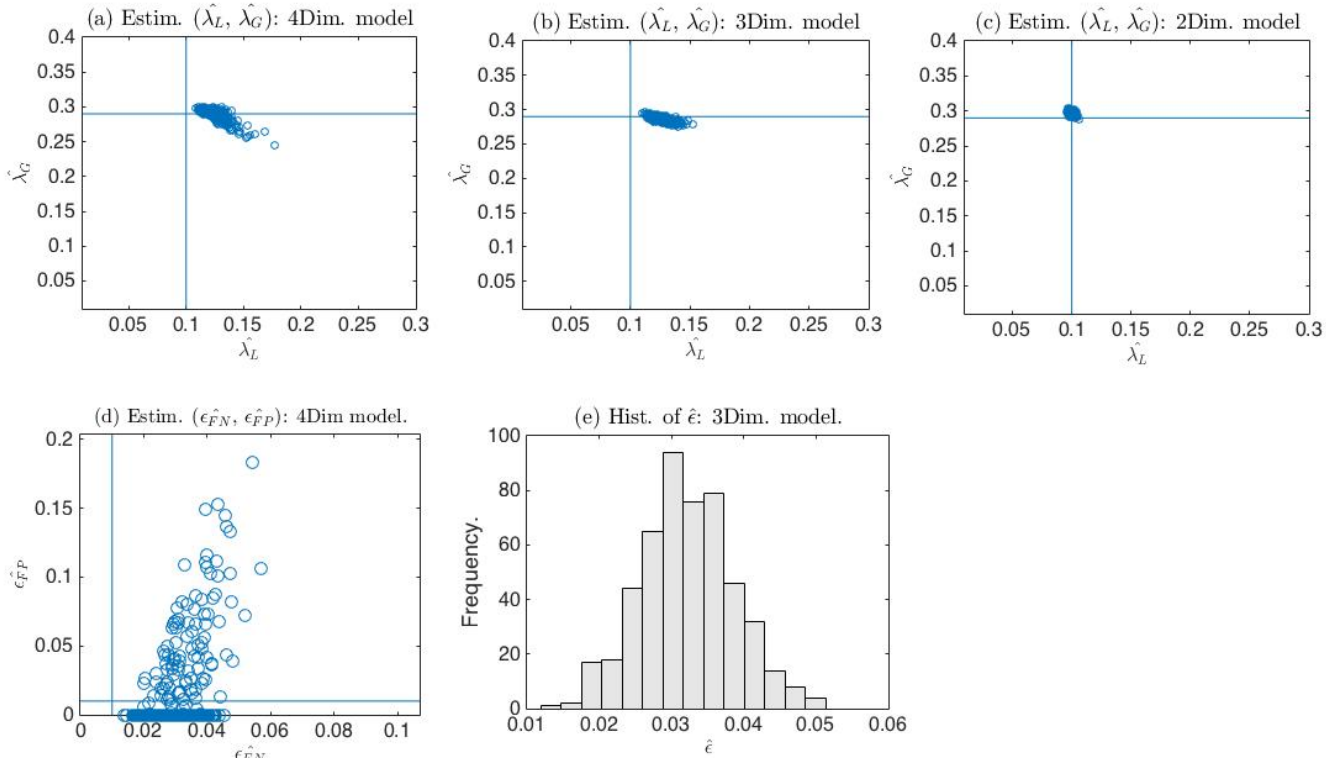

Figure 4: Plots of the estimates with exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution and $\varepsilon = 0.01$.

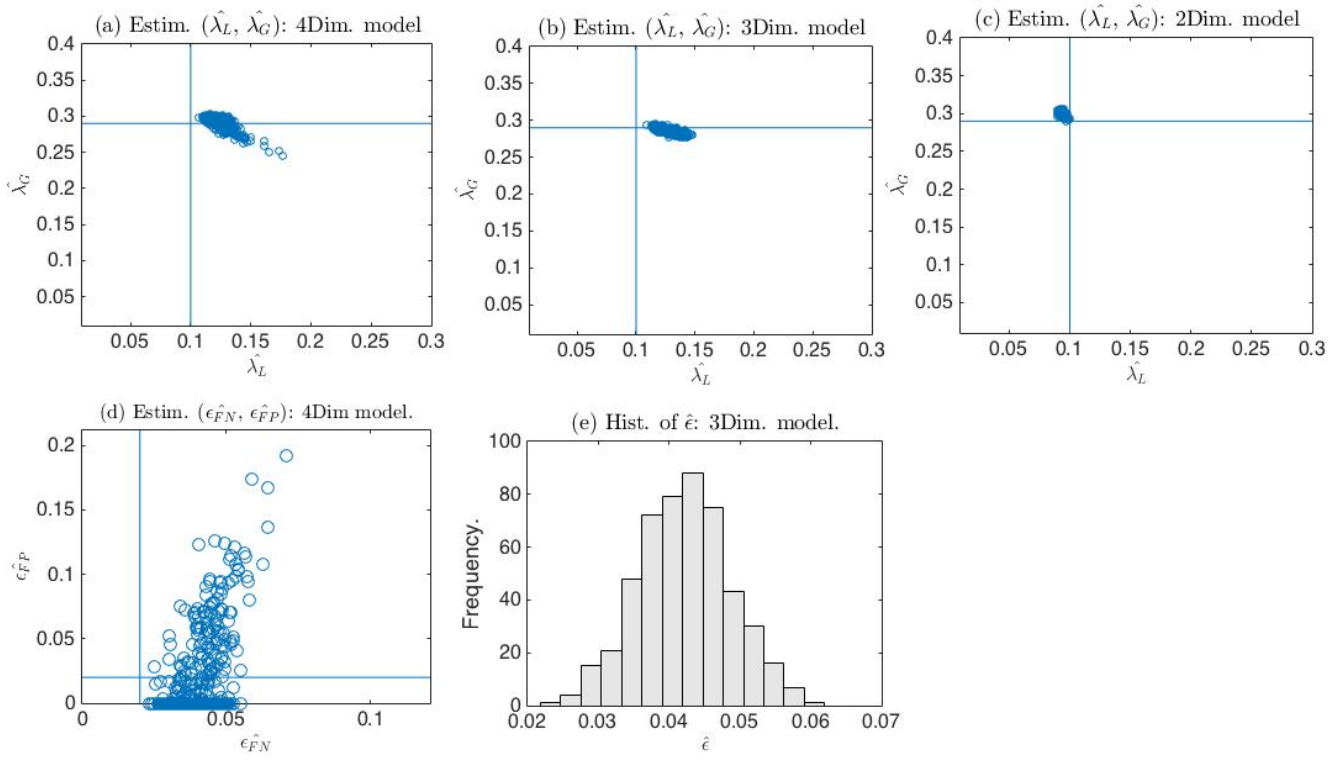

Figure 5: Plots of the estimates with exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution and $\varepsilon = 0.02$.

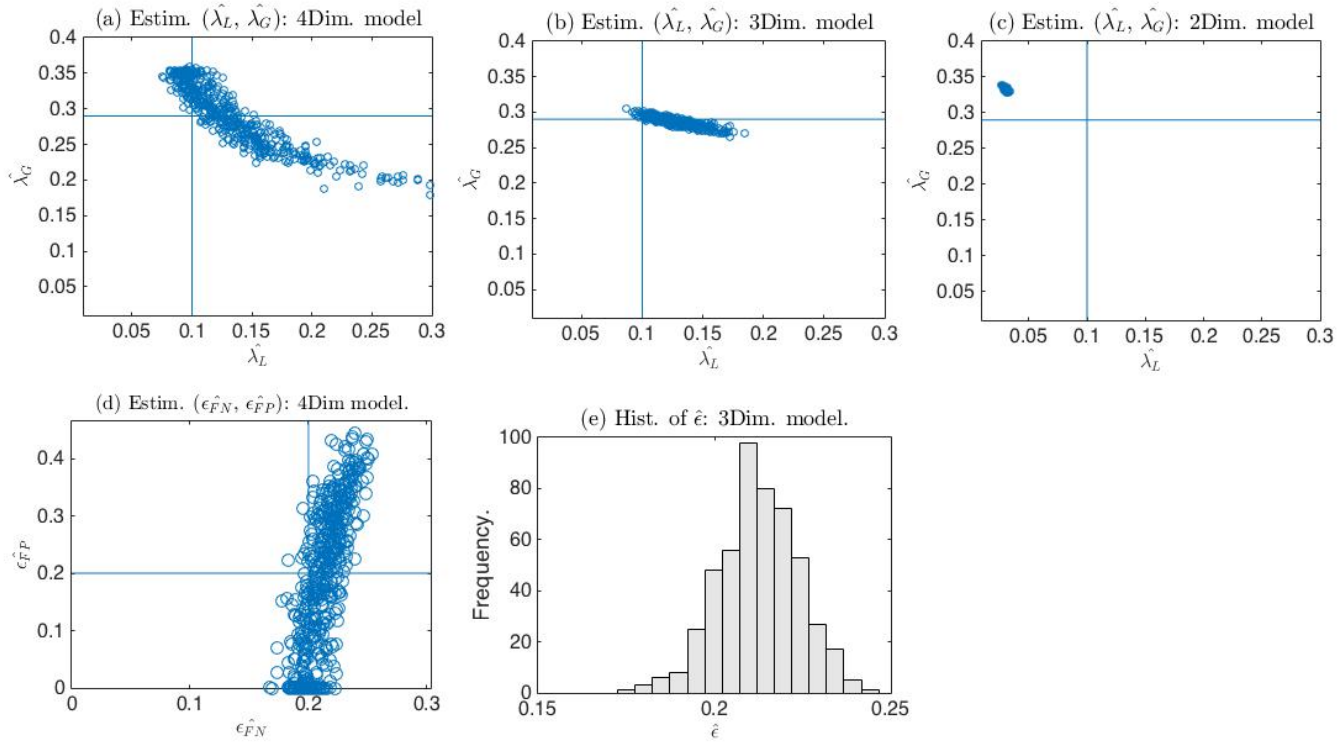

Figure 6: Plots of the estimates with exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution and $\varepsilon = 0.2$.

	Model								
	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim	2Dim	3Dim	4Dim
ε	0.01	0.01	0.01	0.02	0.02	0.02	0.2	0.2	0.2
$\hat{\lambda_L}$	0.0018158	0.028146	0.025563	0.0060862	0.028645	0.025722	0.069539	0.032211	0.082761
$\hat{\lambda_G}$	0.006906	0.0051639	0.025563	0.009607	0.0052508	0.025722	0.043042	0.0072334	0.082761
$\hat{\pi}$	0.015189	0.010833	0.021286	0.016179	0.010735	0.023495	0.01048	0.011482	0.084135
$\hat{z}$	0.013232	0.0294	0.034151	0.008749	0.02967	0.035171	0.074762	0.030281	0.053042
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.021537	N/A	N/A	0.021216	N/A	N/A	0.020137
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.027328	N/A	N/A	0.032086	N/A	N/A	0.12833
$\hat{\varepsilon}$	N/A	0.023163	N/A	N/A	0.023114	N/A	N/A	0.016796	N/A
$\hat{R_*}$	0.053869	0.14262	0.16354	0.02936	0.14404	0.16818	0.38065	0.14746	0.24705

Table 6: Root mean square error of the parameter estimates with exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution.

	$\varepsilon_{FN} = 0.02, \varepsilon_{FP} = 0.1.$			$\varepsilon_{FN} = 0.3, \varepsilon_{FP} = 0.2.$			Theo.
	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.	Param.
$\hat{\lambda_L}$	0.077163	0.077248	0.087328	0.018492	0.12408	0.11088	0.1
$\hat{\lambda_G}$	0.30254	0.3025	0.27016	0.32022	0.23888	0.28331	0.29
$\hat{\pi}$	0.39733	0.39734	0.46066	0.48234	0.53371	0.45364	0.4199
$\hat{z}$	0.7264	0.7265	0.68375	0.55535	0.64279	0.69246	0.7298
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.0068771	N/A	N/A	0.29176	N/A
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.14662	N/A	N/A	0.18414	N/A
$\hat{\varepsilon}$	N/A	0.00019064	N/A	N/A	0.30141	N/A	N/A
$\hat{R_*}$	2.1701	2.1707	2.0246	1.5303	1.8966	2.0626	2.2166

Table 7: Table of mean of the parameter estimates when the epidemic is simulated with exp(4.1) and estimated with Gamma$(2, 4.1/2)$ infectious period distributions.

standard deviation and root mean square error.

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

We simulate epidemic data with Gamma$(2, 4.1/2)$ infectious period distribution and estimate the model with exp(4.1) infectious period distribution. We then present plots of the estimates and table of mean, standard deviation and root mean square error.

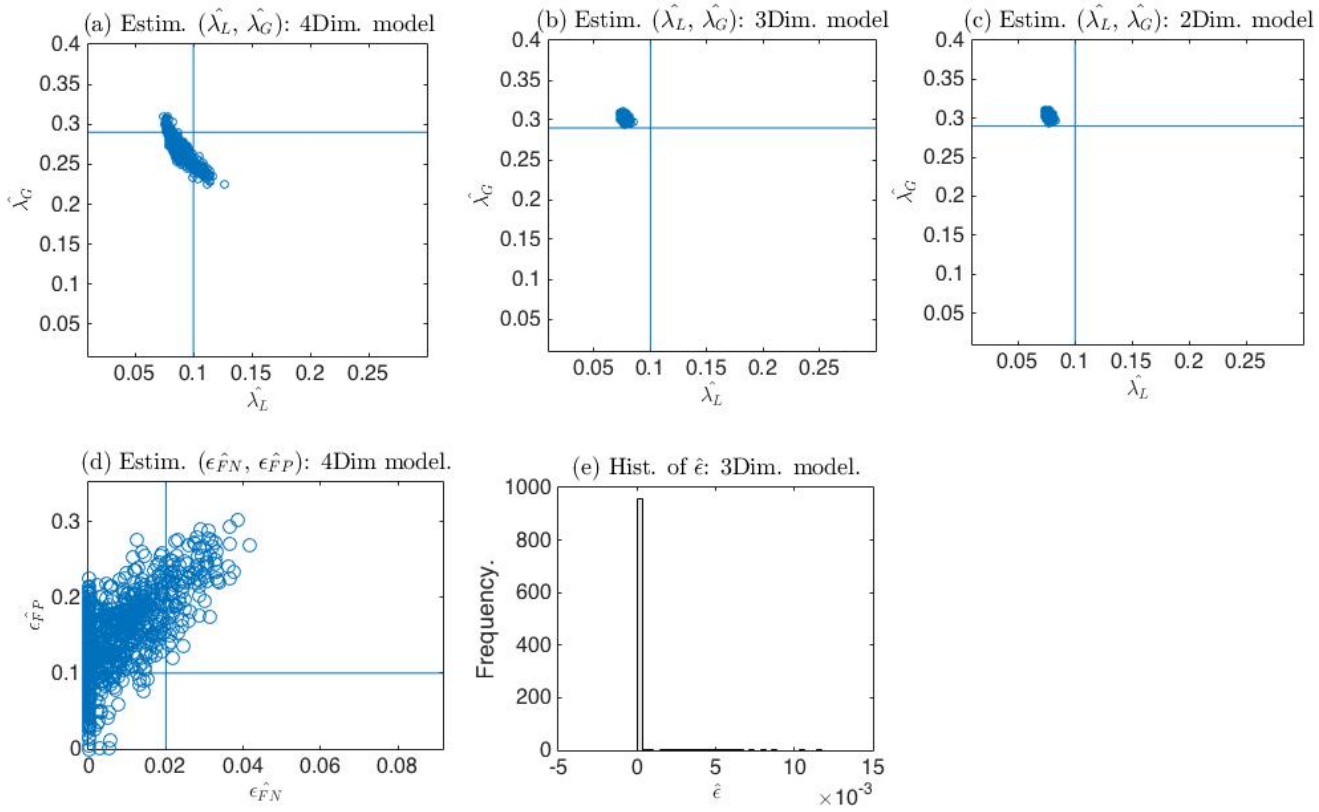

Figure 7: Plots of the estimates using Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution and $\varepsilon_{FN} = 0.02$, $\varepsilon_{FP} = 0.1$.

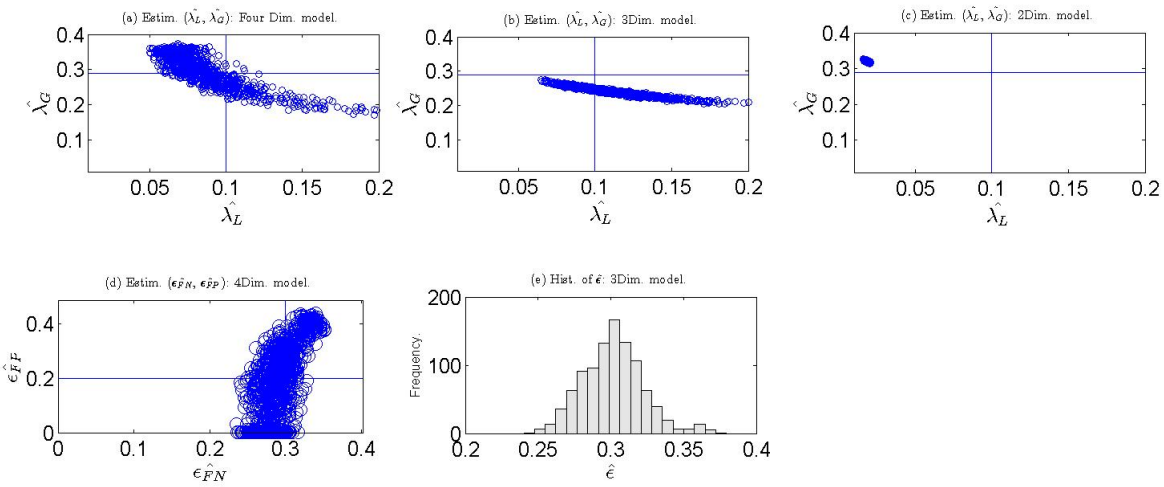

Figure 8: Plots of the estimates using Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution and $\varepsilon_{FN} = 0.3$, $\varepsilon_{FF} = 0.2$.

	$\varepsilon_{FN} = 0.02, \varepsilon_{FP} = 0.1.$			$\varepsilon_{FN} = 0.3, \varepsilon_{FP} = 0.2.$		
	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.
$\hat{\lambda_L}$	0.0013641	0.0014384	0.0075362	0.00083274	0.098363	0.085154
$\hat{\lambda_G}$	0.0025697	0.0025712	0.0025712	0.0018798	0.017461	0.06154
$\hat{\pi}$	0.0050036	0.005004	0.026227	0.003523	0.019542	0.11398
$\hat{z}$	0.0042381	0.0042876	0.017032	0.0029415	0.017353	0.06668
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.0089027	N/A	N/A	0.022678
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.0089027	N/A	N/A	0.022678
$\hat{\varepsilon}$	N/A	0.0010166	N/A	N/A	0.02093	N/A
$\hat{R}_*$	0.018388	0.018763	0.055725	0.0054852	0.051097	0.23005

Table 8: Table of standard deviation the parameter estimates when the epidemic is simulated with exp(4.1) and estimated withGamma$(2, 4.1/2)$ infectious period distributions.

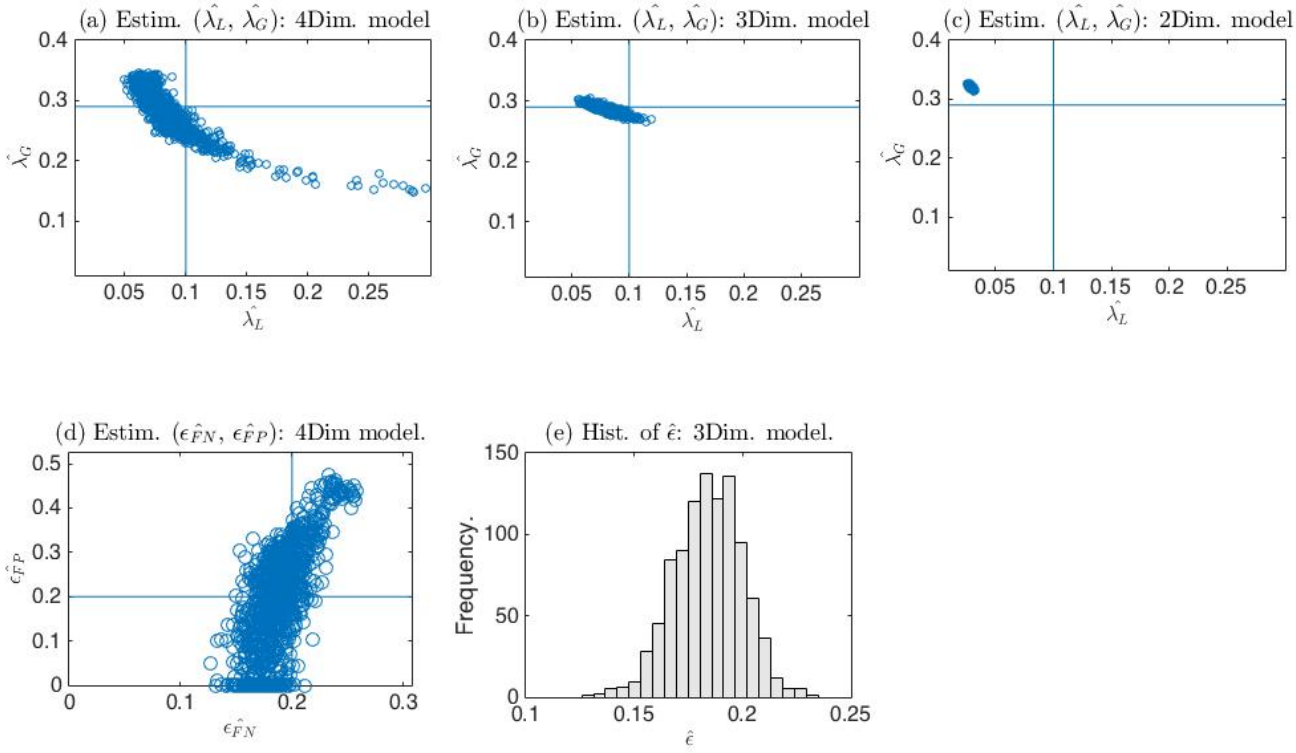

Figure 9: Plots of the estimates using Gamma$(2, 4.1/2)$ infectious period distribution when the epidemic data is simulated with exp(4.1) infectious period distribution and $\varepsilon_{FN} = 0.2$, $\varepsilon_{FF} = 0.2$.

	$\varepsilon_{FN} = 0.02, \varepsilon_{FP} = 0.1.$			$\varepsilon_{FN} = 0.3, \varepsilon_{FP} = 0.2.$		
	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.
$\hat{\lambda_L}$	0.022877	0.022797	0.014742	0.081513	0.10122	0.085804
$\hat{\lambda_G}$	0.012802	0.01276	0.023996	0.030277	0.054021	0.085804
$\hat{\pi}$	0.0077887	0.0077836	0.063068	0.062538	0.11547	0.11881
$\hat{z}$	0.019558	0.019481	0.064056	0.17446	0.088712	0.076388
$\hat{\varepsilon_{FN}}$	N/A	N/A	0.015855	N/A	N/A	0.024118
$\hat{\varepsilon_{FP}}$	N/A	N/A	0.070549	N/A	N/A	0.13442
$\hat{\varepsilon}$	N/A	0.059818	N/A	N/A	0.0555	N/A
$\hat{R_*}$	0.1225	0.12201	0.27237	0.68633	0.32402	0.27674

Table 9: Table of root mean square error of the parameter estimates when the epidemic is simulated with exp(4.1) and estimated with Gamma$(2, 4.1/2)$ infectious period distributions.

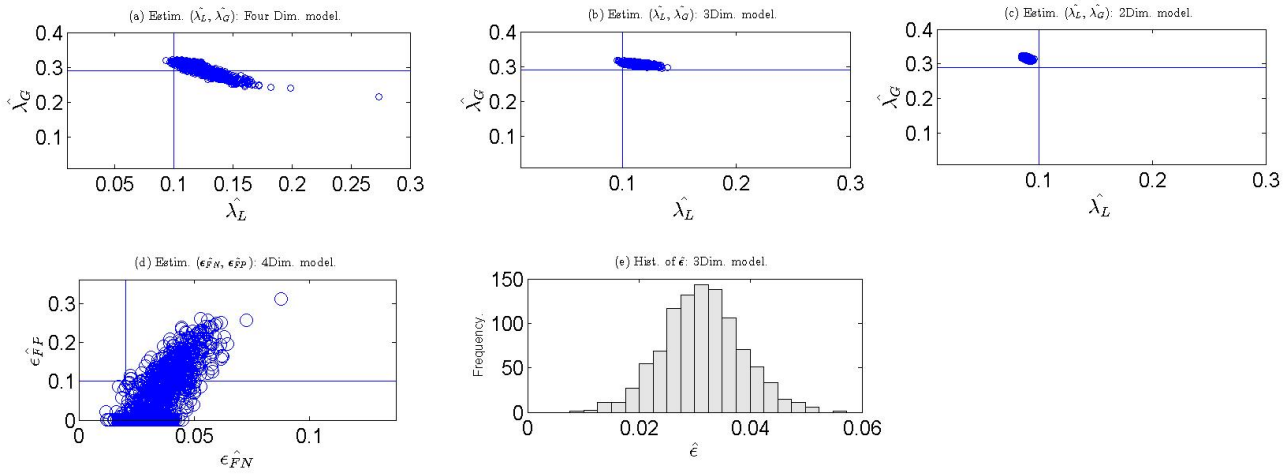

Figure 10: Plots of the estimates using exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution and $\varepsilon_{FN} = 0.02$, $\varepsilon_{FP} = 0.1$.

4. Results and Discussion.

In figures 1 (a)-(e), the estimates are biased and imprecise with less variability from the two and three dimensional models as shown in figures 1 (b) and (c) owing to misspecification. The three dimensional model is better than the two and four dimensional models.

Similar pattern of behaviours in figures 1 (a)-(c) can be seen in figures 2 (a)-(c). The three dimensional model is better than the two and four dimensional models.

In figures 3 (a)-(e), we see large variability of the estimates of the four dimensional model around their true values compared to those of the three dimensional model. While those of the two dimensional model are biased and imprecise. In general the three dimensional model is better than the two and four dimensional models.

In figure 8 (c), the estimates of the two dimensional model are more precise with less variability than the three and four dimensional models in figures 8 (a) and (b) owing to misspecification.

Similar behaviours in figures 8 (a)-(e) can be seen in figures 9 (a)-(e).

In figures 6 (a)-(e), the estimates of the three and four dimensional models are centered at their true values with more variability for the four dimensional model than those of the three dimensional model. While those of the two dimensional model are imprecise and biased.

In general the three dimensional model is better than the two and four dimensional models

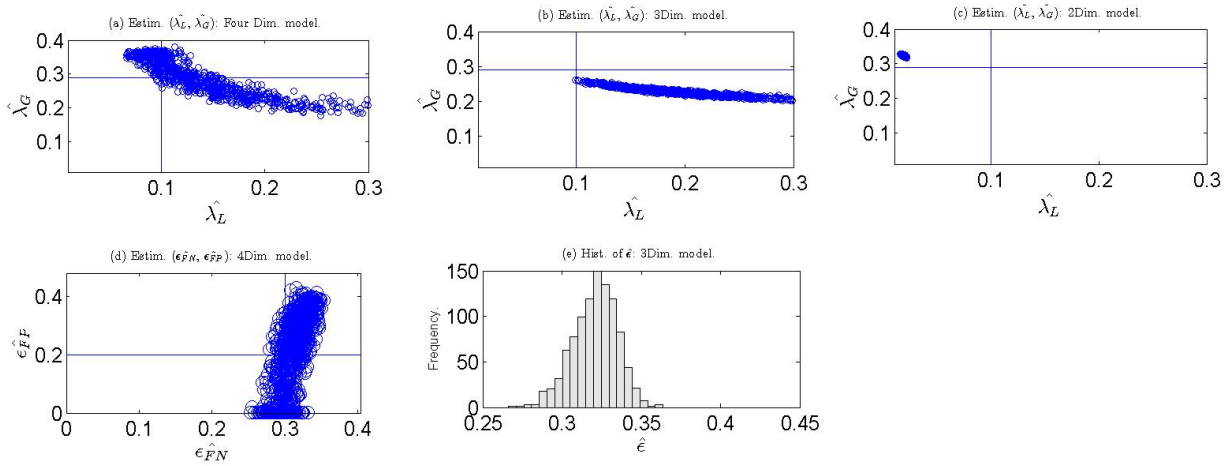

Figure 11: Plots of the estimates using exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution and $\varepsilon_{FN} = 0.3$, $\varepsilon_{FP} = 0.2$.

	$\varepsilon_{FN} = 0.02, \varepsilon_{FP} = 0.1.$			$\varepsilon_{FN} = 0.3, \varepsilon_{FP} = 0.2.$			$\varepsilon_{FN} = 0.2, \varepsilon_{FP} = 0.2.$			Theo.
	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.	Param.
$\hat{\lambda_L}$	0.0899	0.11325	0.12146	0.018857	0.20678	0.17525	0.030476	0.12636	0.14328	0.1
$\hat{\lambda_G}$	0.31604	0.3072	0.29637	0.32328	0.22549	0.28042	0.33295	0.28666	0.28834	0.29
$\hat{\pi}$	0.38268	0.38487	0.403	0.47353	0.53291	0.43831	0.41922	0.41955	0.42131	0.4291
$\hat{z}$	0.74134	0.75816	0.74916	0.564	0.68191	0.7314	0.63686	0.73928	0.74019	0.7117
$\varepsilon_{\hat{FN}}$	N/A	N/A	0.035793	N/A	N/A	0.31001	N/A	N/A	0.21176	N/A
$\varepsilon_{\hat{FP}}$	N/A	N/A	0.073987	N/A	N/A	0.19614	N/A	N/A	0.18777	N/A
$\hat{\varepsilon}$	N/A	0.031352	N/A	N/A	0.32081	N/A	N/A	0.21156	N/A	N/A
$\hat{R_*}$	2.2213	2.3271	2.2885	1.5441	2.0035	2.221	1.7299	2.2436	2.2554	2.1106

Table 10: Table of mean of the parameter estimates when the epidemic is simulated with Gamma$(2, 4.1/2)$ and estimated with exp(4.1) infectious period distributions..

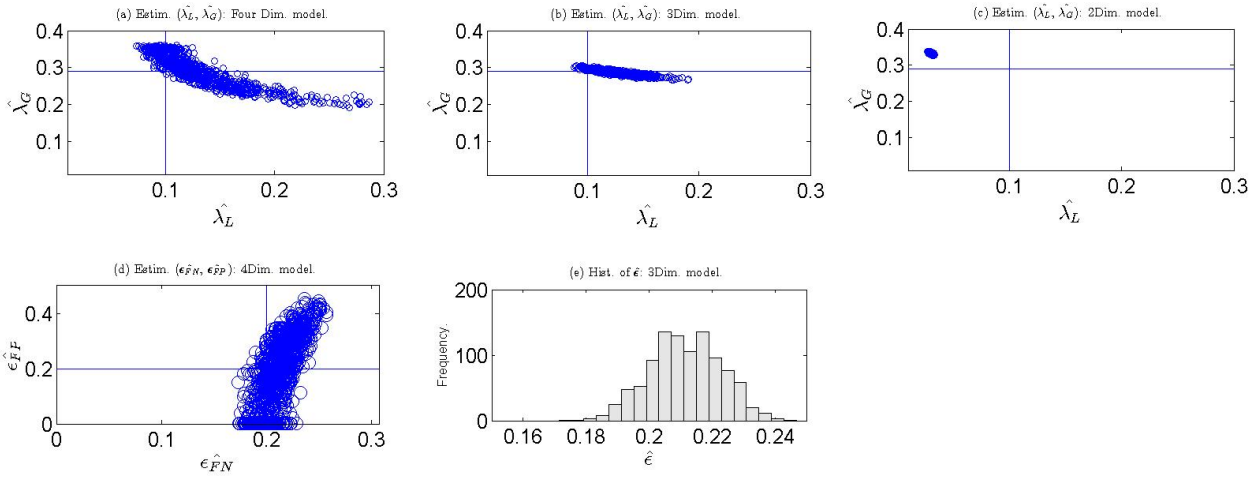

Figure 12: Plots of the estimates using exp(4.1) infectious period distribution when the epidemic data is simulated with Gamma$(2, 4.1/2)$ infectious period distribution and $\varepsilon_{FN} = 0.2$, $\varepsilon_{FP} = 0.2$.

	$\varepsilon_{FN} = 0.02, \varepsilon_{FP} = 0.1.$			$\varepsilon_{FN} = 0.3, \varepsilon_{FP} = 0.2.$			$\varepsilon_{FN} = 0.2, \varepsilon_{FP} = 0.2.$		
	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.
$\hat{\lambda_L}$	0.0016924	0.0066428	0.01531	0.00088043	0.10474	0.33149	0.0010178	0.016721	0.079516
$\hat{\lambda_G}$	0.0025711	0.0033824	0.017484	0.0018071	0.013616	0.060984	0.0020738	0.0064708	0.048072
$\hat{\pi}$	0.0044885	0.0047762	0.029084	0.0031539	0.016093	0.10675	0.0035174	0.0069734	0.083025
$\hat{z}$	0.0034881	0.0055627	0.015519	0.0025597	0.014854	0.057308	0.0027811	0.010356	0.043984
$\hat{\varepsilon_{FN}}$	N/A	N/A	0.0096162	N/A	N/A	0.017639	N/A	N/A	0.01552
$\hat{\varepsilon_{FP}}$	N/A	N/A	0.0096162	N/A	N/A	0.013944	N/A	N/A	0.01552
$\hat{\varepsilon}$	N/A	0.0070962	N/A	N/A	0.32081	N/A	N/A	0.19768	N/A
$\hat{R_*}$	0.015789	0.032068	0.070022	0.0051863	0.042476	0.24985	0.0071748	0.05349	0.19768

Table 11: Table of standard deviation the parameter estimates when the epidemic is simulated with Gamma$(2, 4.1/2)$ and estimated with exp(4.1) infectious period distributions.

	$\varepsilon_{FN}=0.02, \varepsilon_{FP}=0.1.$			$\varepsilon_{FN}=0.3, \varepsilon_{FP}=0.2.$			$\varepsilon_{FN}=0.2, \varepsilon_{FP}=0.2.$		
	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.	2Dim.	3Dim.	4Dim.
$\hat{\lambda_L}$	0.01024	0.014822	0.026358	0.081148	0.14954	0.33976	0.069531	0.031216	0.090497
$\hat{\lambda_G}$	0.026164	0.01753	0.018601	0.033325	0.065926	0.061702	0.043003	0.00728	0.048076
$\hat{\pi}$	0.046586	0.044434	0.01860	0.044594	0.1051	0.1071	0.010441	0.011785	0.083343
$\hat{z}$	0.029886	0.046832	0.04058	0.14768	0.033251	0.060585	0.07485	0.029491	0.052406
$\hat{\varepsilon_{FN}}$	N/A	N/A	0.018487	N/A	N/A	0.0202759	N/A	N/A	0.019465
$\hat{\varepsilon_{FP}}$	N/A	N/A	0.073775	N/A	N/A	0.13298	N/A	N/A	0.12738
$\hat{\varepsilon}$	N/A	0.029513	N/A	N/A	0.072166	N/A	N/A	0.016198	N/A
$\hat{R_*}$	0.11175	0.21884	0.19116	0.56651	0.11525	0.27301	0.3808	0.14333	0.24492

Table 12: Table of root mean square error of the parameter estimates when the epidemic is simulated with Gamma$(2, 4.1/2)$ and estimated with exp(4.1) infectious period distributions.

With misclassification error in the data and misspecification, the estimates of the three dimensional model are biased with less variability around their true values than those of the four dimensional model. Those of the two dimensional model are biased and imprecise.

In general, the three dimensional model is better than the two and four dimensional models on three dimensional epidemic data with model misspecification.

In figures 7 (a)-(c), the estimates of the three models are biased and imprecise.

In figures 8 (a) and (b), we see large variability of the estimates of the three and four dimensional models around their true values. While those of the two dimensional model are biased and imprecise. The three and four dimensional models are better than the two dimensional model.

In figures 9 (a) and (b), the scatter plots of the estimates of the four and three dimensional models are centered around their true values but with more variability from those of the four dimensional model. While those of the two dimensional model are biased and imprecise. Given this scenario, the three dimensional model is significantly better than the two and four dimensional models as theoretically expected.

In figures 10 (a)-(c), the scatter plots of the estimates of λ_L and λ_G from the three and four dimensional models are close to their true values with more variability from those of the four dimensional model. While those of the two dimensional model are biased.

In figures 11, (a)-(e), similar behaviours in figures 10 (a)-(c) are shown with less variability.

Also, similar behaviours in figures 9 (a)-(d) are repeated in figures 12 (a)-(d)

From the scatter plots (a)-(e) in figures 7 , 8 and 9 and tables 7, 8 and 9, we see that estimates of the four dimensional model are more precise than those of the two and three dimensional models in the face of misspecification when the epidemic data is four dimensional data. While in figures 10-12, (a)-(e), with misspecification the three and four dimensional models are better than the two dimensional models, when the data are both misclassified and misspecified.

5 Conclusion.

From the scatter plots (a)-(e) in figures 7 , 8 and 9 and tables 7, 8 and 9, we see that estimates of the four dimensional model are more precise than those of the two and three dimensional models in the face of misspecification when the epidemic data is four dimensional data. While in figures 10-12, (a)-(e), with misspecification and misclassification of the final size data, the three and four dimensional models are give better estimates than the two dimensional models. These behaviours are similiar to those observed in [16], iin which with misspecification of the infectious period distribution, the estimates are biased, with model that does not fit well to the final data, which is indicative of the model behaviour with misclassified final data. Thus, further studies are often required to identify whether these behaviours are attributable to misspecification of the infectious period distribution or misclassification of the final size epidemic data. in model estimation.

Acknowledgments

The Author will like to acknowledge Dr. Owen D. Lyne, and Professor Martin Ridout for their valuable contributions.

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, **47**(3) : 961-974, (1991).

[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, **20**(2) : 227-241, (1983).

[4] D. Clancy and P. D. O'NEILL, Exact Bayesian Inference and Model Selection for Stochastic Models of Epidemics Among a Community of Households. Scandinavian Journal of Statistics, **34**(2) : 259-274, (2007).

[5] 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, **18**(2) : 289-310, (1986).

[6] F. BALL, A Note on the Total Size Distribution of Epidemic Models. Journal of Applied Probability, **23**(3) : 832-836, (1986).

[7] F. G. BALL, D. MOLLISON AND G. SCALIA-TOMBA, Epdemics with Two Levels of Mixing. Annals of Applied Probability, **7**(1) : 46-89, (1997).

[8] 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, **126** : 115-125, (2000).

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

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

[11] P. GUSTAFSON, Measurement error and Misclassification in Statistics and Epidemiology, Impacts and Bayesian Adjustment. Chapman and Hall/ CRC, (2009).

[12] M. E. HALLORAN, I. M. LONGINI AND C. J. STRUCHINER, Design and Analysis of Vaccine Studies. Statistics for Biology and Health, Springer, (2010).

[13] I. M. LONGINI, JR AND J.S. KOOPMAN, Household and Community Transmission Parameters from Final Distribution of Infections in Households. Biometrics, **38**(1) :115-126, (1982).

[14] 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**(5) : 736-750, (1982)

[15] 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, **16** : 3093-3108, (2020)

[16] A. M. UMAR, Stochastic Household epidemic Model with Misspecification. Journal of progressive research in Mathematics, **18**(3) : 31-46, (2021)

[17] A. M. UMAR, Stochastic SIR Household Epidemic Model with Misclassification. Unpulished, (2017).