

RESEARCH ARTICLE

Open Access

# Estimating the basic reproduction number for single-strain dengue fever epidemics

Adnan Khan<sup>1</sup>, Muhammad Hassan<sup>2</sup> and Mudassar Imran<sup>1\*</sup>

## Abstract

**Background:** Dengue, an infectious tropical disease, has recently emerged as one of the most important mosquito-borne viral diseases in the world. We perform a retrospective analysis of the 2011 dengue fever epidemic in Pakistan in order to assess the transmissibility of the disease. We obtain estimates of the basic reproduction number  $R_0$  from epidemic data using different methodologies applied to different epidemic models in order to evaluate the robustness of our estimate.

**Results:** We first estimate model parameters by fitting a deterministic ODE vector-host model for the transmission dynamics of single-strain dengue to the epidemic data, using both a basic ordinary least squares (OLS) as well as a generalized least squares (GLS) scheme. Moreover, we perform the same analysis for a direct-transmission ODE model, thereby allowing us to compare our results across different models. In addition, we formulate a direct-transmission stochastic model for the transmission dynamics of dengue and obtain parameter estimates for the stochastic model using Markov chain Monte Carlo (MCMC) methods. In each of the cases we have considered, the estimate for the basic reproduction number  $R_0$  is initially greater than unity leading to an epidemic outbreak. However, control measures implemented several weeks after the initial outbreak successfully reduce  $R_0$  to less than unity, thus resulting in disease elimination. Furthermore, it is observed that there is strong agreement in our estimates for the pre-control value of  $R_0$ , both across different methodologies as well across different models. However, there are also significant differences between our estimates for the post-control value of the basic reproduction number across the two different models.

**Conclusion:** In conclusion, we have obtained robust estimates for the value of the basic reproduction number  $R_0$  associated with the 2011 dengue fever epidemic before the implementation of public health control measures. Furthermore, we have shown that there is close agreement between our estimates for the post-control value of  $R_0$  across the different methodologies. Nevertheless, there are also significant differences between the estimates for the post-control value of  $R_0$  across the two different models.

**Keywords:** Epidemiology, Dengue fever, Statistical inference, Stochastic model, Markov chain Monte Carlo

## Multilingual abstracts

Please see Additional file 1 for translations of the abstract into the six official working languages of the United Nations.

## Background

Global incidence of dengue has seen a striking increase over the past few decades [1,2]. The infectious disease is now endemic in more than a hundred tropical and

subtropical countries worldwide [1-3]. With an estimated 50–100 million cases and nearly 10,000–20,000 deaths annually, dengue ranks second to Malaria amongst deadly mosquito-borne diseases [1,2,4-6]. The disease is caused by one of four virus serotypes (strains) of the genus *Flavivirus* [2,3,7]. Most infected individuals suffer from dengue fever, a severe flu-like illness characterized by high fever, which is not usually a threat to mortality [2,8]. The symptoms of the disease last one to two weeks, after an initial incubation period of about 4–7 days [9]. Some infected individuals however, develop dengue hemorrhagic fever (DHF) resulting in bleeding, low levels of blood platelets and blood plasma leakage, or dengue shock

\*Correspondence: mimran@asu.edu

<sup>1</sup>Department of Mathematics, Lahore University of Management Sciences, DHA, Lahore, Pakistan

Full list of author information is available at the end of the article

syndrome (DSS) resulting in extremely low blood pressures. The risk associated with DHF and DSS is considerably higher, with mortality ranging from 5–15% [3,5,9,10].

There is evidence of dengue epidemics occurring in North America, Asia and Africa in the late 18<sup>th</sup> century [3]. Up until the middle of the 20<sup>th</sup> century however, incidences of dengue fever have been rare [3]. Nonetheless, since the 1970's, there has been a marked increase in the number of dengue cases, as well as the frequency and severity dengue epidemics, with the WHO claiming a 30-fold increase in the incidence of dengue between 1960 and 2010 [2,3,8]. Factors such as population growth, rapid urbanization and increase international travel are often cited as having contributed to this dramatic increase [8]. Dengue is currently endemic in nearly 110 countries in Southeast Asia, the Americas, Africa and the Eastern Mediterranean [2]. The WHO estimates that nearly 2.5 billion people are at risk of contracting the disease. Furthermore, nearly 50–100 million cases and almost 20,000 deaths due to more severe forms of dengue fever are reported globally every year, making dengue one of the deadliest mosquito-transmitted diseases [1,2,4-6].

Dengue is transmitted to humans through mosquito bites. Female mosquitos of the *Aedes* genus, primarily *Aedes aegypti*, acquire the dengue virus through a blood meal from infected humans [2,11]. The dengue virus has an incubation period of about 7–10 days in the vector, and is then spread to susceptible humans who are bitten by the infected mosquito [9]. The virus also has an incubation period of 4–7 days in the host [9]. While vectors never recover from infection with the dengue virus, the infection in hosts lasts only about one to two weeks [2]. Hosts that recover from infection with one serotype of the dengue virus gain life-long immunity from that serotype but only temporary and partial immunity to other serotypes [2,4,12-14]. This partial cross-immunity is the cause of antibody-dependent enhancement (ADE) in the setting of a secondary infection with a different serotype of DENV (Dengue Virus). ADE is hypothesized to be one factors leading to DHF and DSS, the more severe form of dengue disease [4,12,13,15]. In this study however, we will consider infection involving only a single serotype of the dengue virus.

About 80% of individuals suffering from a primary infection with DENV are asymptomatic or display only a mild, uncomplicated fever [2,8]. A much smaller proportion of infected individuals suffer from dengue hemorrhagic fever and dengue shock syndrome [3,5,9]. As mentioned previously however, risk of DHF and DSS is associated primarily with secondary infection with a heterologous serotype of DENV [4,12,13,15]. In general, the course of infection of dengue can be divided into three separate phases: febrile, critical and recovery. The febrile phase, which is rarely life threatening, is marked by the sudden onset of high fever,

rash, headaches and muscle and joint pains, which lead to the alternative name "breakbone fever" for dengue disease [8]. While most individuals then progress to the recovery phase, a small fraction of infected individuals instead progress to the critical phase of the disease. This phase lasts for one or two days and is marked by low blood pressure, leakage of blood plasma from the capillaries and decreased blood supply to organs. Severe cases of these symptoms are associated with DHF and DSS and the mortality in this phase of the disease is estimated to be as high as 5–15% [3,5,8,9].

Over the past several years, a number of deterministic mathematical models have been proposed to analyze the transmission dynamics of dengue in urban communities [5,11-17]. L. Esteva and C. Vargas [14] have investigated the coexistence of two serotypes of dengue virus using a deterministic ODE model. Moreover, Ferguson et al. [15] have investigated the effects of ADE on the transmission of multiple serotypes of dengue virus. In addition, Garba et al. [11] have shown the existence of a backward bifurcation in a standard incidence ODE model for a single strain of dengue virus. Garba et al. [12] have also explored the effects of cross-immunity on the transmission dynamics of two strains of dengue virus. Similarly, H. Wearing and P. Rohani [13] have investigated the effects of both ADE and cross immunity on multiple serotypes of dengue virus. Finally, Chowell et al. [18] have estimated the basic reproduction number for dengue using spatial epidemic data.

In addition, over the past few decades, several stochastic epidemic models for the spread of infectious diseases have also been proposed and analyzed [19-27]. An important qualitative difference between deterministic and stochastic epidemic models in general is the asymptotic dynamics [28]. Furthermore, stochastic models also allow for the possibility of disease extinction in finite time and therefore the expected time to disease extinction can be calculated [19,28,29]. It is also observed that stochastic models better capture the uncertainty and variability that is inherent in real-life epidemics due to factors such as the unpredictability of person-to-person contact [27,29]. L. J. S. Allen [28,29] has explored the utility of stochastic epidemic models by comparing them with deterministic models. Despite, the utility of stochastic models, however, very little stochastic modeling has been performed for the transmission dynamics of dengue virus (see [26] and the references therein).

The purpose of this study is to estimate the transmissibility of the dengue virus during the 2011 dengue fever epidemic in Pakistan using epidemic data in the form of the cumulative number of reported cases of dengue. We will employ three different techniques, applied to two different models and compare the results across both the different statistical inference methodologies as well

as the different models. The first approach, based on the earlier work of Cintron-Arias et al. [30], will involve fitting a deterministic epidemic model for the transmission of dengue to the epidemic data using an ordinary least squares (OLS) scheme implemented using the built-in optimization toolbox in MATLAB and applied in the context of an appropriate statistical model. The second method, also based on the recent work of Cintron-Arias et al. [30], will use a generalized least squares (GLS) scheme to fit the same deterministic model to the epidemic data. Furthermore, both approaches will also be applied to a different direct-transmission model for the transmission dynamics of dengue. Finally, the third approach will involve the formulation of a direct transmission stochastic epidemic model for dengue. We shall then use Markov chain Monte Carlo techniques to obtain a probability distribution for the model parameters.

A simple but effective measure of the transmissibility of an infectious disease is given by the basic reproduction number  $R_0$ , defined as the total number of secondary infections produced by introducing a single infective in a completely susceptible population [31]. For vector-borne diseases such as malaria and dengue,  $R_0$  is the number of secondary cases produced by a single infectious vector introduced in a completely susceptible host and vector population. In general, for simple epidemic models, if  $R_0$  is greater than unity, an epidemic will occur while if  $R_0$  is less than unity, an outbreak will most likely not occur. Thus, the value of  $R_0$  can be used to determine the intensity of control measures that need to be implemented in order to contain the epidemic.

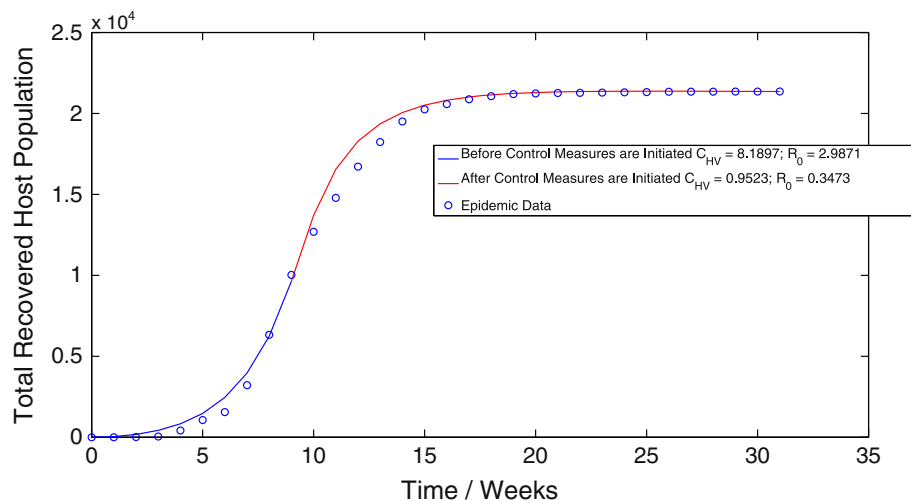
The estimation of the basic reproductive number is generally an indirect process because the model parameters that  $R_0$  depends on are difficult or impossible to determine directly. The general methodology used

therefore, attempts to fit an epidemic model to available epidemic data in order to estimate the model parameters. These parameters are then used to estimate the basic reproduction number  $R_0$ . The current study is based on this methodology.

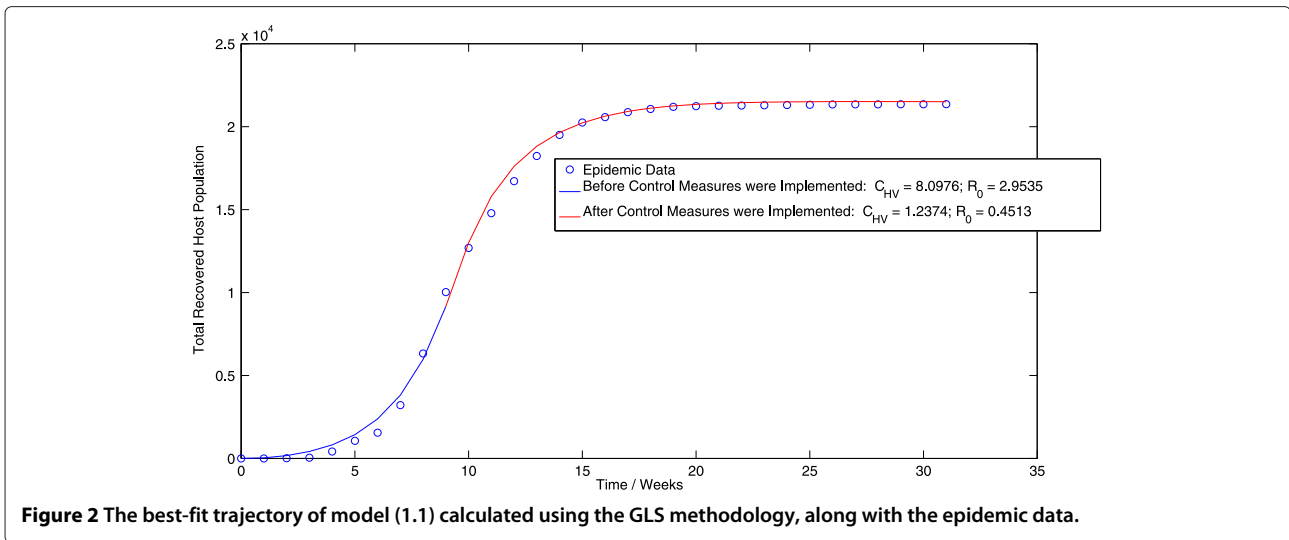
## Results

Applying the algorithm for the ordinary least squares (OLS) methodology to the vector-host model (1.1) results in a value of  $C_{HV} = 8.1897 \text{ week}^{-1}$  before the implementation of control measures and a value of  $C_{HV} = 0.9523 \text{ week}^{-1}$  after the implementation of control measures. Thus, we obtain an estimate of  $R_0 = 2.9871$  before the implementation of control measures and  $R_0 = 0.3473$  after the implementation of control measures. The best-fit trajectory of model (1.1) calculated using the OLS methodology, along with the epidemic data is displayed in Figure 1. Similarly, implementing the algorithm for the generalized least squares (GLS) methodology results in a value of  $C_{HV} = 8.0976 \text{ week}^{-1}$  before the implementation of control measures and a value of  $C_{HV} = 1.2374 \text{ week}^{-1}$  after the implementation of control measures. Thus, when using the GLS scheme, we obtain an estimate of  $R_0 = 2.9535$  before the implementation of control measures and  $R_0 = 0.4513$  after the implementation of control measures. Figure 2 displays the best-fit trajectory of model (1.1) calculated using the GLS methodology. We observe that the GLS estimates for  $R_0$  before the implementation of control measures are in close agreement with the results obtained using the OLS scheme. There is however, difference between the estimates for  $R_0$  after the implementation of control measures.

Application of the algorithm for the ordinary least squares (OLS) methodology to the direct-transmission



**Figure 1** The best-fit trajectory of model (1.1) calculated using the OLS methodology, along with the epidemic data.

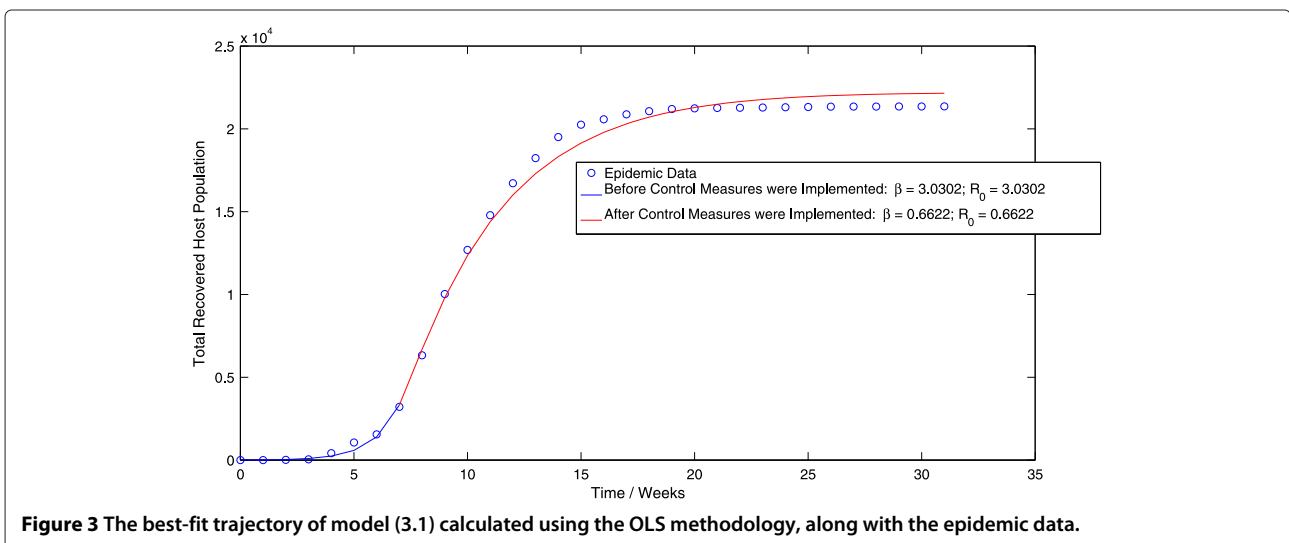


**Figure 2** The best-fit trajectory of model (1.1) calculated using the GLS methodology, along with the epidemic data.

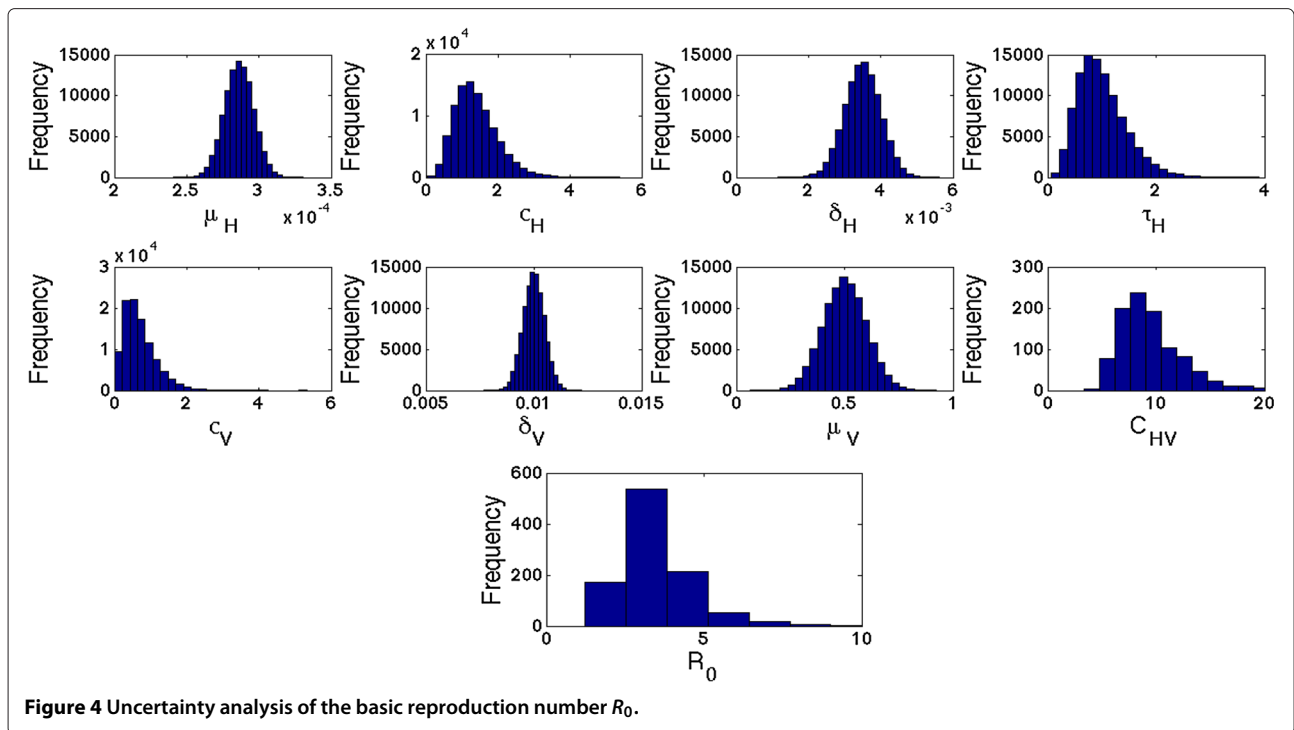
model (3.1) results in a value of  $\beta = 3.0302 \text{ week}^{-1}$  before the implementation of control measures and a value of  $\beta = 0.6622 \text{ week}^{-1}$  after the implementation of control measures. This corresponds to a value of  $R_0 = 3.0182$  before the implementation of control measures and  $R_0 = 0.6596$  after the implementation of control measures. Figure 3 displays the best-fit trajectory of the direct-transmission model (3.1) calculated using the OLS scheme. We observe that the pre-control value of  $R_0$  is slightly larger (1.04%) than the corresponding value obtained by applying the OLS scheme to the vector-host model (1.1), which we feel is not a statistically significant increase. However, the post-control value of  $R_0$  is significantly higher (89.92%) than the corresponding value obtained by applying the OLS scheme to the vector-host model (1.1).

In addition, pre-control estimates of  $R_0$  for the dengue epidemic from the uncertainty analysis are shown in Figure 4. The 95% confidence interval for the pre-control value of  $R_0$  is given by (2.0293, 6.5310). The probability that  $R_0$ , before the implementation of control measures, is greater than unity is more than 99.9%. Similarly, the sensitivity analysis, shown in Figure 5, suggests that the most significant (PRCC values above 0.5 or below  $-0.5$ ) sensitivity parameters to  $R_0$  are  $\tau_H$ ,  $\sigma_V$  and  $\mu_V$ . This suggests that these parameters need to be estimated with precision in order to accurately capture the transmission dynamics of dengue.

Furthermore, implementing the algorithm for the generalized least squares methodology in the case of the direct-transmission model (3.1) results in a value of  $\beta = 3.0920 \text{ week}^{-1}$  before the implementation of control measures



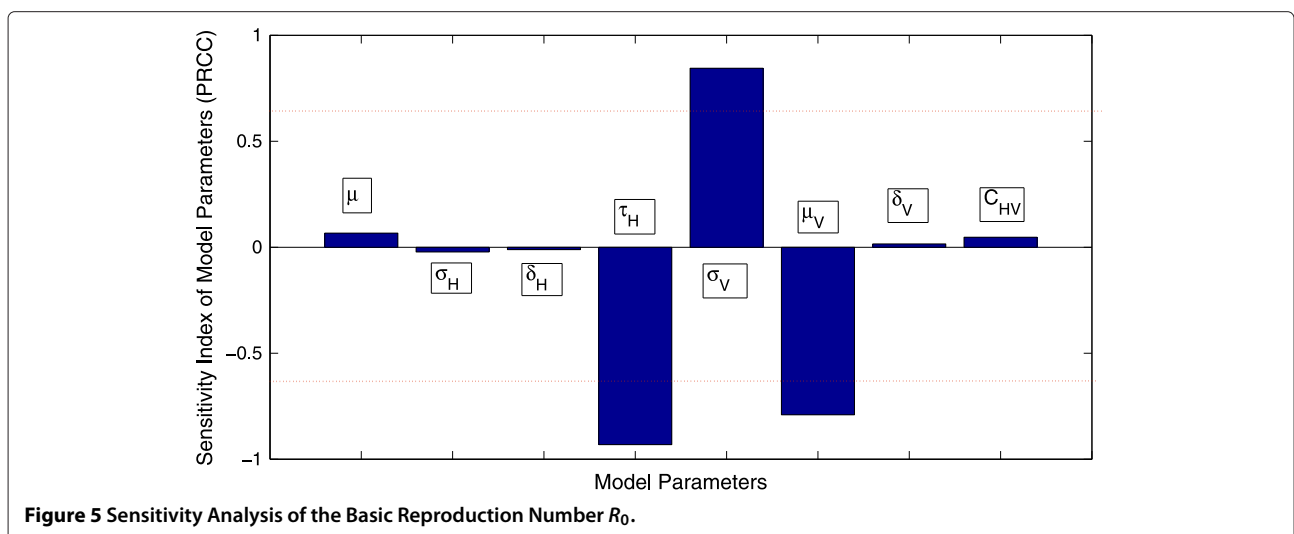
**Figure 3** The best-fit trajectory of model (3.1) calculated using the OLS methodology, along with the epidemic data.

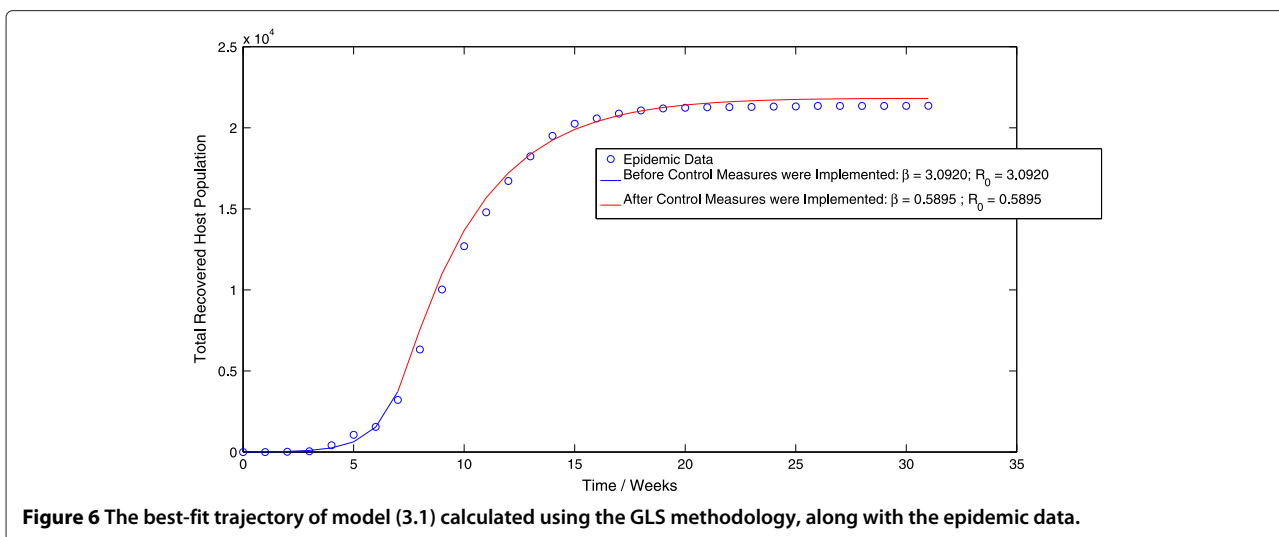


and a value of  $\beta = 0.5895 \text{ week}^{-1}$  after the implementation of control measures. These values of the contact rate result in an estimate of  $R_0 = 3.0797$  before the implementation of control measures and  $R_0 = 0.5872$  after the implementation of control measures. The best-fit trajectory of the direct-transmission model (3.1) calculated using the GLS scheme is displayed in Figure 6. We observe that while these results are in close agreement with the corresponding results obtained by applying the OLS scheme to the direct-transmission SEIR model, the estimated value of post-control  $R_0$  is again, significantly

larger (30.11%) than the corresponding value obtained by applying the GLS scheme to the vector-host model. However, there is only a slight increase (4.27%) in the estimate for the pre-control value of  $R_0$  across the different models, which we again feel is not a statistically significant increase.

Finally, implementation of the MCMC algorithm to the stochastic direct-transmission model (4.1) results in a value of  $\beta = 3.0650 \text{ week}^{-1}$  before the implementation of control measures and a value of  $\beta = 0.6318 \text{ week}^{-1}$  after the implementation of control measures. These values of



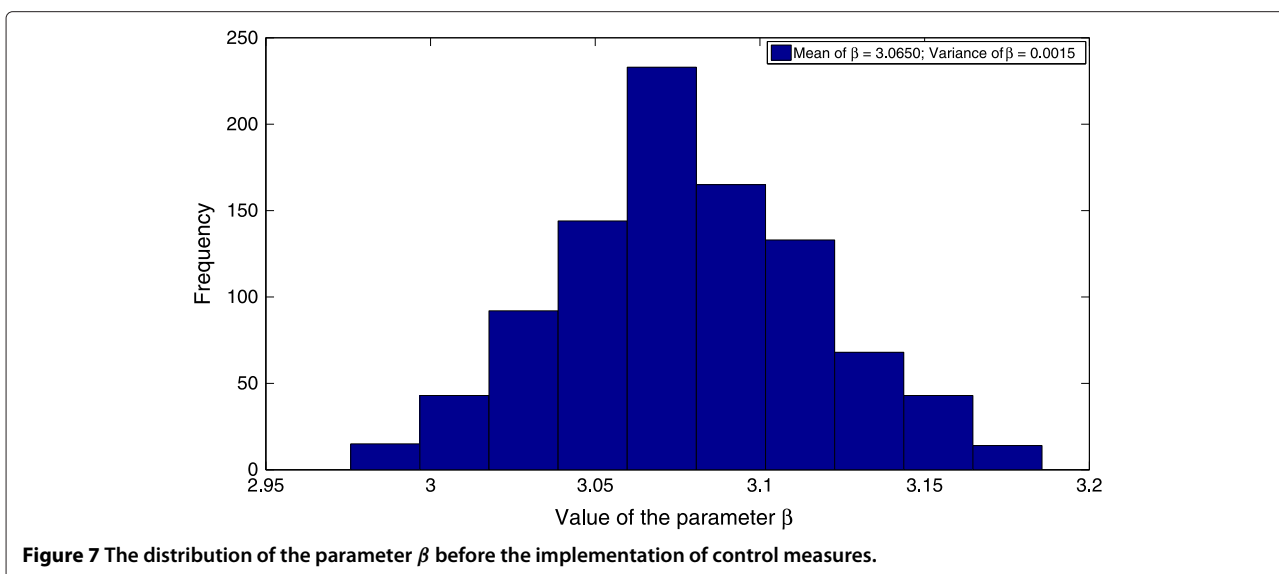


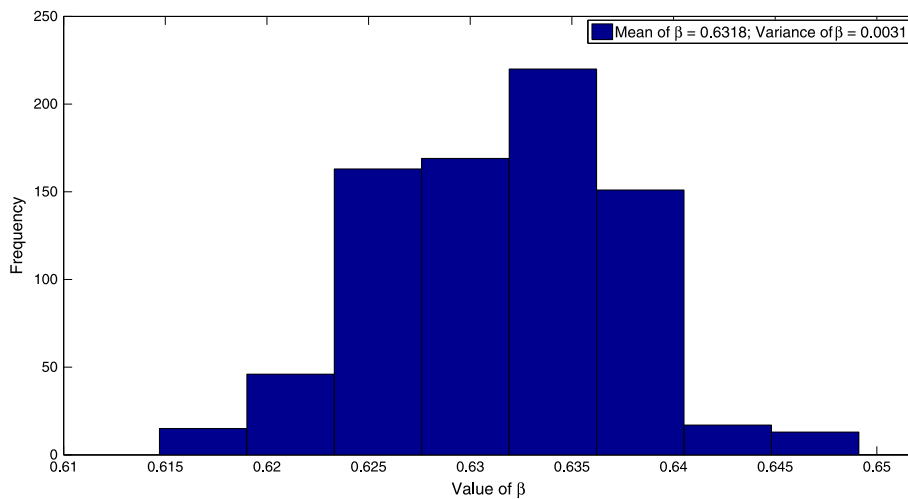
the contact rate correspond to values of  $R_0 = 3.0528$  before the implementation of control measures and  $R_0 = 0.6293$  after the implementation of control measures. The probability distributions of the contact rate  $\beta$ , both before and after the implementation of control measures are displayed in Figure 7 and Figure 8 respectively. Moreover, the mean trajectory of the stochastic model (4.1) calculated using Monte Carlo simulations involving the mean of the posterior distribution is displayed in Figure 9. The results of the MCMC methodology are in close agreement with the results obtained by applying the GLS and OLS schemes to the deterministic direct-transmission model (3.1). Furthermore, there is also close agreement in our estimates for the pre-control value of  $R_0$  across both the previous methodologies as well as the different models. There is however, again, a significant difference between

the post-control value of  $R_0$ , obtained using the MCMC algorithm and the estimates of the post-control value of  $R_0$ , obtained by applying the OLS and GLS scheme to the vector-host model (1.1).

### Discussion

We have performed a retrospective analysis of the 2011 dengue fever epidemic in Pakistan and obtained estimates of the basic reproduction number  $R_0$ , from epidemic data using three different methods.  $R_0$ , defined as the total number of secondary infections produced by introducing a single infective in a completely susceptible population, is a simple but effective measure of the transmissibility of an infectious disease. In each case it was observed that the value of  $R_0$  was initially well in excess of unity, leading to the observed epidemic outbreak. Some





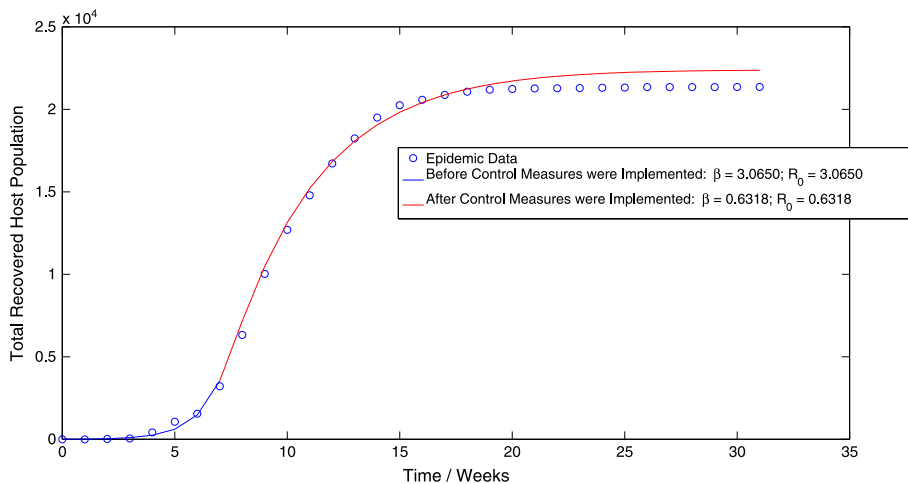
**Figure 8** The distribution of the parameter  $\beta$  after the implementation of control measures.

weeks after the initial outbreak however, control measures were successfully implemented that reduced the value of  $R_0$  to less than unity, thus resulting in disease elimination.

Several methods have been proposed for the estimation of  $R_0$ , both for deterministic as well as for stochastic models. These methods depend upon the mathematical model of the disease as well the nature of the data. In the case of deterministic compartmental models, least squares fit to the data has been widely used to estimate the model parameters [18,30,32]. For stochastic models likelihood based techniques have been used by several authors [33,34]. We consider two ODE based deterministic models and a Continuous Time Markov Chain (CTMC) based stochastic model. Using least squares estimation for the

deterministic models and a Markov Chain Monte Carlo based approach for the stochastic model, we compare the value of  $R_0$  obtained using the different models. We note that this is the first such study performed for the Dengue Epidemic in Pakistan.

The first inference methodology involved fitting a deterministic ODE model for the transmission dynamics of single-strain dengue to the epidemic data using a basic ordinary least squares (OLS) scheme in the context of a statistical model which assumed longitudinally constant variance for the epidemic data. An a priori more realistic methodology was used to fit the deterministic ODE model to obtain estimates of the model parameters using a generalized least squares (GLS) scheme which made use of a statistical model that assumed that variances associated



**Figure 9** The mean trajectory of the stochastic model (4.1), calculated using Monte Carlo simulations involving the mean of the posterior distribution, plotted along with the epidemic data.

with the observation process were directly proportional to the measurement values. One of the questions we tried to address was whether or not the variances were strongly dependent on the observations. Finally, we formulated a discrete-time, direct transmission, stochastic model for the spread of dengue virus and used Markov chain Monte Carlo (MCMC) methods to perform Bayesian inference and estimate the basic reproduction number.

We observe that the estimates for the basic reproduction number  $R_0$  before the implementation of control measures are in excellent agreement for the same model across different methodologies. Similarly, across different models there is a very slight but nonetheless statistically insignificant difference in our estimates of the pre-control basic reproduction number. We therefore conclude that our estimation of the pre-control value of  $R_0$  is quite robust, both across different methodologies as well as across different models. This leads us to believe that the noise does not depend significantly on the data. Furthermore, agreement in our estimates across models indicates that both the vector-host model as well as the direct-transmission model can be used to accurately capture the disease dynamics of actual dengue epidemics before the implementation of control measures.

While there is also close agreement in our estimates for the basic reproduction number  $R_0$  after the implementation of control measures across different methodologies, there is nonetheless significant difference between the post-control estimates of  $R_0$  across the vector-host and direct-transmission models. Specifically,  $R_0$  is estimated to be significantly larger in value when using the direct-transmission model as opposed to the vector-host model. We conjecture that this might be due to the fact that the direct-transmission model makes use of an approximation, which involves solving for the equilibrium value of the vector force of infection. Thus, while the vector force of infection rises and peaks for the vector-host model, before settling to its equilibrium value, it is in effect equal to the smaller equilibrium value for the direct-transmission model. Therefore, since the vector force of infection for the direct transmission model is, in effect, smaller for the time period after the implementation of control measures, it results in a larger estimate of the basic reproduction number in order to produce a 'best-fit' for the observed epidemic data.

## Conclusion

In conclusion, we have attempted to assess the transmissibility of the dengue virus during the 2011 dengue fever epidemic in Pakistan by estimating the basic reproduction number  $R_0$  both before and after the implementation of public health control measures. Our estimates for the

pre-control value of  $R_0$  are in close agreement both across different methodologies and the different models. Furthermore, the post-control estimates are also in close agreement across the different methodologies. There is however, a significant increase in the estimates of the post-control value of  $R_0$  obtained using the direct-transmission model compared to estimates obtained using the vector-host model.

## Methods

### Methods and materials for statistical inference using the vector-host model

#### *The vector-host epidemic Model*

The model is a deterministic vector-host ODE model that assumes a homogeneous mixing of the host (human) and vector (mosquito) populations. The total human population at time  $t$ , denoted by  $N(t)$ , is divided into four mutually exclusive classes comprising of susceptible humans  $S_H(t)$ , exposed humans  $E_H(t)$ , infected humans  $I_H(t)$  and recovered humans  $R_H(t)$ . It is assumed that individuals who recover from infection with a particular serotype of Dengue gain lifelong immunity to it [2]. Similarly, the total vector population at time  $t$  is denoted by  $N_V(t)$  and is divided into three mutually exclusive classes comprising susceptible of susceptible vectors  $S_V(t)$ , exposed vectors  $E_V(t)$  and infected vectors  $I_V(t)$ . It is assumed that vectors (mosquitoes) infected with a particular serotype of Dengue never recover [2]. We also modify the original model of Garba et al. [11] by assuming that exposed humans and exposed vectors do not transmit the disease.

The model assumes that the susceptible human population  $S_H(t)$  has a constant recruitment rate  $\Pi_H$  and natural death rate  $\mu$ . Susceptible individuals are infected with Dengue virus (due to contact with infected vectors) at a rate  $\lambda_H$  and thus enter the exposed class  $E_H$ . The exposed population  $E_H(t)$  is depleted at the natural death rate  $\mu$ . Additionally, exposed individuals develop symptoms and move into the infected class  $I_H$  at a rate  $\sigma_H$ . The infected population  $I_H(t)$  is depleted via the natural death rate  $\mu$ , the disease-induced death rate  $\delta_H$  and the recovery rate of infected individuals  $\tau_H$ . Finally, the recovered population  $R_H(t)$  decreases due to the natural death rate  $\mu$ .

Similarly, the susceptible vector population  $S_V(t)$  has a constant recruitment rate  $\Pi_V$  and a natural death rate  $\mu_V$ . Susceptible vectors are infected with Dengue virus (due to effective contact with infected humans) at a rate  $\lambda_V$  and thus move to the exposed vector class  $E_V$ . The exposed vector class  $E_V(t)$  is depleted at the natural death rate  $\mu_V$ . In addition exposed vectors develop symptoms and move to the infected vector class  $I_V$  at a rate  $\sigma_V$ . Infected vectors, in addition to the natural death rate  $\mu$  die at a disease induced death rate  $\delta_V$ .



Mathematically, the model is as follows

$$\begin{aligned}
 \frac{dS_H}{dt} &= \Pi_H - \lambda_H S_H - \mu S \\
 \frac{dE_H}{dt} &= \lambda_H S - (\sigma_H + \mu) E_H \\
 \frac{dI_H}{dt} &= \sigma_H E_H - (\tau_H + \mu + \delta_H) I_H \\
 \frac{dR_H}{dt} &= \tau_H I_H - \mu R_H \\
 \frac{dS_V}{dt} &= \Pi_V - \lambda_V S_V - \mu_V S_V \\
 \frac{dE_V}{dt} &= \lambda_V S_V - (\sigma_V + \mu_V) E_V \\
 \frac{dI_V}{dt} &= \sigma_V E_V - (\mu_V + \delta_V) I_V
 \end{aligned} \tag{1.1}$$

where,

$$\begin{aligned}
 \lambda_H &= C_{HV} \frac{I_V}{N_H} \\
 \lambda_V &= C_{HV} \frac{I_H}{N_H}
 \end{aligned}$$

A description of the variables and parameters of the model (1.1) is given in Table 1.

Garba et al. [11] have calculated the basic reproduction number  $R_0$  for their original model. Although, we have made a slight modification to the original model, the basic reproduction number for our model is not significantly different. Hence, for the model (1.1),  $R_0$  is given by

$$R_0 = \sqrt{\frac{C_{HV}^2 \Pi_V \mu \sigma_H \sigma_V}{\Pi_H \mu_V k_1 k_2 k_3 k_4}} \tag{1.2}$$

where

$$\begin{aligned}
 k_1 &= \sigma_H + \mu_H & k_2 &= \tau_H + \mu_H + \delta_H \\
 k_3 &= \sigma_V + \mu_V & k_4 &= \mu_V + \delta_V
 \end{aligned}$$

**Table 1 Description of the variables of the vector-host model (1.1)**

Variable	Description
$N_H(t)$	Total host population
$S_H(t)$	Population of susceptible hosts
$E_H(t)$	Population of exposed hosts
$I_H(t)$	Population of infected hosts
$R_H(t)$	Population of recovered hosts
$N_V(t)$	Total vector population
$S_V(t)$	Population of susceptible vectors
$E_V(t)$	Population of exposed vectors
$I_V(t)$	Population of infected vectors

Representative trajectories of the model (1.1) for different values of  $R_0$  are given in Figure 10 and Figure 11. As expected, a value of  $R_0$  greater than unity leads to an epidemic while a value of  $R_0$  less than unity leads to swift disease elimination.

#### Data sources

The epidemic data used in this study was collected by the Punjab Disaster Management Authority (PDMA) during the 2011 Dengue Epidemic in Punjab, Pakistan. The data, displayed in Figure 12, represents the cumulative number of dengue fever cases reported over a 32-week period extending from the 1<sup>st</sup> of August 2011, to the 20<sup>th</sup> of February 2012. The data was collected from a number of public and private hospitals in the major metropolitan centers of Punjab, including Lahore, Faisalabad and Multan. Patients were classified as infected with dengue virus based on the results of laboratory tests for the dengue specific antibodies Immunoglobulin G (IgG) and Immunoglobulin M (IgM). As per the Government of Pakistan's directives, the laboratory test was available at a subsidized rate in all major hospitals in the capital city of Lahore. It is therefore unlikely that poverty played a serious role in under-reporting of dengue fever cases.

Prior to August 2011, there were three reported cases of dengue fever, all occurring several months before the actual epidemic. This leads us to conclude that these were isolated incidents and were not directly related to the epidemic itself. Furthermore, nearly 87% of all dengue infections were caused by a single serotype (DEN2) of the dengue virus. This justifies our use of a single-strain epidemic model as opposed to dengue models that incorporate the effects of cross-immunity and ADE.

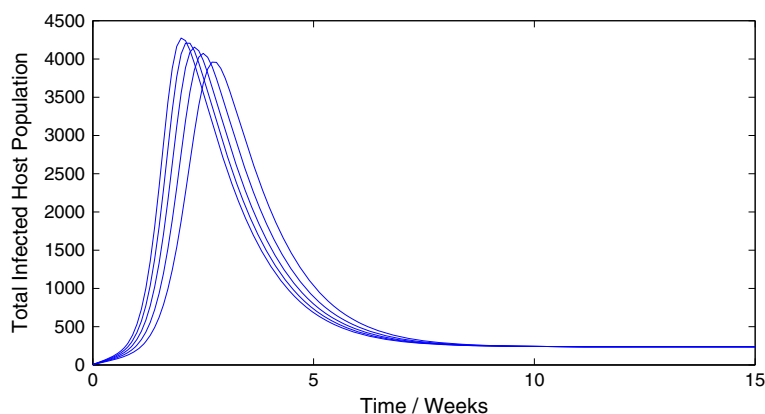
#### Estimation schemes

In order to calculate  $R_0$ , we require the values of several parameters used in model (1.1). Furthermore, we require knowledge of the initial conditions that will be used to simulate trajectories of the model (1.1).

Estimates for several of the model parameters used in model (1.1) can be obtained from existing studies on Dengue Fever. Table 2 lists these parameters along with reasonable estimates of their values.

The recruitment rate for the susceptible host population depends on the demographics of the urban population that is being considered. This study uses epidemic data collected during the 2011 Dengue Epidemic in Punjab, Pakistan. Therefore, in the absence of concrete estimates, the host recruitment rate has been chosen so as to allow for a realistic steady state host population.

The effective contact rate  $C_{HV}$ , which is a measure of the rate at which contact between an infective and a susceptible individual occurs and the probability that such contact will lead to an infection, is extremely difficult to determine



**Figure 10** Sample trajectories of model (1.1) with  $R_0 > 1$ .

directly. Most previous studies have used assumed values for the effective contact rate [11-13]. Thus, it is not possible to directly estimate the basic reproduction number. Consequently, we will adopt an indirect approach, similar to previous studies such as [30] and [18], by first finding the value of the parameter  $C_{HV}$  for which the model (1.1) has the best agreement with the epidemic data, and then using the resultant parameter values to estimate  $R_0$ .

As mentioned before, for the purpose of simulating model (1.1) we require knowledge of the initial conditions. It is possible to consider the initial conditions  $(S_H(0), E_H(0), I_H(0), R_H(0), S_V(0), E_V(0), I_V(0))$  as model parameters along with the effective contact rate  $C_{HV}$  and estimate values for all parameters. Such a technique, however, produces slightly unreliable results. This is explained by the fact that the available epidemic data is restricted to the cumulative number of dengue cases reported, while the optimization schemes that we will employ produces estimates for eight variables. There are thus too many degrees of freedom and the ‘best-fit’ may result in unrealistic estimates for the initial conditions.

We will therefore use reasonable estimates for the initial conditions and restrict ourselves to optimizing only the effective contact rate  $C_{HV}$ .

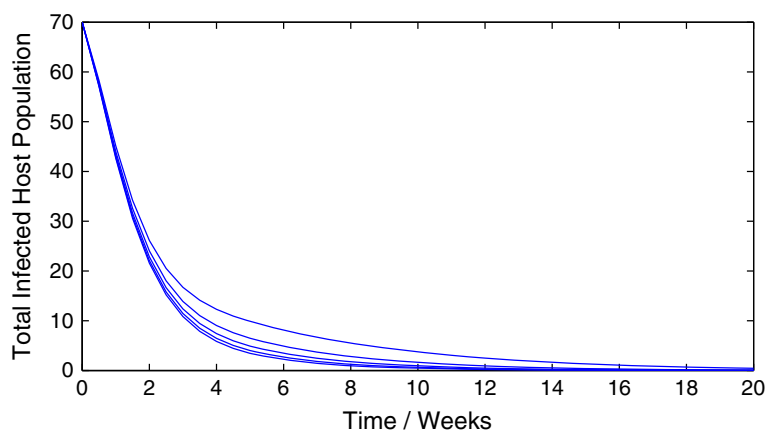
Table 3 displays the initial conditions that were therefore chosen.

In the following sections we will employ different methods to estimate the parameter  $\omega = C_{HV}$  by minimizing the difference between the predictions of model (1.1) and the epidemic data.

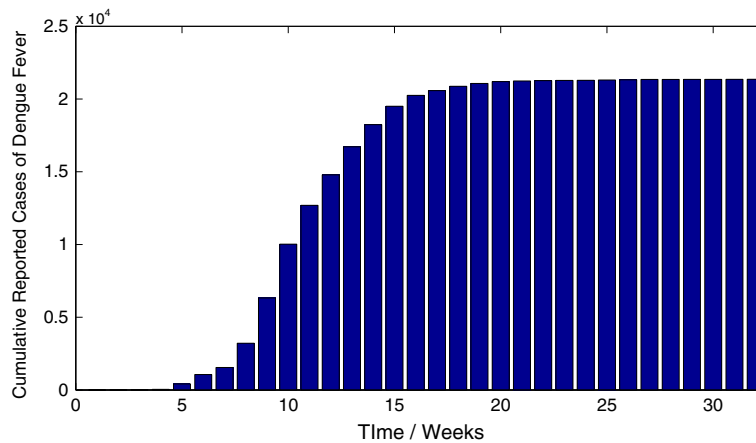
#### Ordinary Least Squares Estimation

We will first attempt to estimate the effective contact rate by fitting the best trajectory of model (1.1) to the epidemic data using an ordinary least squares (OLS) scheme, implemented using the *fminsearch* function in the built-in Optimization Toolbox in MATLAB. This will allow us to estimate the parameter  $\omega$  and thus calculate the basic reproduction number  $R_0$ .

In order to fit model trajectories with the observed epidemic data, we will assume that all reported cases recovered from the infection after a time lag of two weeks.



**Figure 11** Sample trajectories of model (1.1) with  $R_0 < 1$ .



**Figure 12** The cumulative number of dengue cases during the 2011 dengue fever epidemic in Pakistan as reported by the Punjab Disaster Management Authority (PDMA).

Thus, the epidemic data, after a lapse of two weeks, represents the total recovered host population. In addition, in order to account for the effect of control measures that were implemented during the actual epidemic, we have assumed that the transmission rate  $C_{HV}$  is a function of time  $t$ . We will assume that the transmission rate was constant up until the point when control measures were implemented, whereupon it changed to a different constant value. An alternative definition of the contact rate is also given in [33]. Thus, mathematically,

$$C_{HV}(t) = \begin{cases} C_{HV1} & t < t^* \\ C_{HV2} & t \geq t^* \end{cases} \quad (2.1)$$

where  $t^*$  is the time at which control measures were first implemented. For the purpose of this study and in view of

no concrete information being available in this regard, we have assumed that  $t^* = 8$  weeks.

For the purpose of this section we shall employ the notation and methodology developed in [30]. Essentially, we will employ inverse problem methodology to obtain estimates of  $\omega = C_{HV}$  by minimizing the difference between the observed weekly cumulative number of recovered host individuals and the model predictions using an ordinary least squares (OLS) criterion. This will be done in the context of an appropriate statistical model.

The ordinary least squares estimation methodology that we will employ is based on model (1.1) as well as an appropriate statistical model for the observation process. Thus, similar to [30], we assume that the model (1.1), together with a ‘true’ value of the parameter  $\omega$ , given by  $\omega_0$ , perfectly describes the transmission dynamics of the dengue epidemic. Moreover, we assume that the  $N$  observations,  $Y_j$ , given by the epidemic data, are generated by a statistical process. However, the  $N$  observations also contain random deviations from the underlying statistical process. Hence, following [30], it is assumed that

$$Y_j = p(t_j, \omega_0) + \epsilon_j \quad \text{for } j = 1, 2, 3, \dots, N \quad (2.2)$$

**Table 2** Description of the parameters of the vector-host model (1.1)

Parameter	Description	Value	Reference
$\Pi_H$	Host recruitment rate	140 week <sup>-1</sup>	assumed
$\Pi_V$	Vector recruitment rate	28000 week <sup>-1</sup>	[12]
$\frac{1}{\mu}$	Host mortality	3494 weeks	[12]
$\delta_H$	Host disease-induced death rate	0.0035 week <sup>-1</sup>	[12]
$\frac{1}{\mu_V}$	Vector mortality	2 weeks	[12]
$\delta_V$	Vector disease-induced death rate	negligible week <sup>-1</sup>	[11]
$\frac{1}{\sigma_H}$	Latency period for exposed hosts	1 week [13]	
$\frac{1}{\tau_H}$	Recovery time for infected hosts	1 week	[12]
$\frac{1}{\sigma_V}$	Latency period for exposed vectors	$\frac{10}{7}$ weeks	[13]
$C_{HV}$	Effective contact rate	Variable	

**Table 3** Initial conditions used when applying the OLS and GLS methodology to the vector-host model (1.1)

Initial condition	Value
$S_H(0)$	1 million
$E_H(0)$	15
$I_H(0)$	3
$R_H(0)$	0
$S_V(0)$	0.1 million
$E_V(0)$	60
$I_V(0)$	20

where  $p(t_j, \omega_0)$  denotes the total number of infected individuals who recover in the time span of week  $j$  when using the ‘true’ parameter  $\omega_0$ . Thus, it is assumed that the observed epidemic data is a particular realization of the statistical model (2.2). Under the framework of model (1.1) therefore,

$$p(t_j, \omega_0) = \int_{t_{j-1}}^{t_j} \tau_H I_H(t, \omega_0) \quad \text{for } j = 1, 2, 3, \dots, N \quad (2.3)$$

where  $t_0$  denotes the time of start of the statistical process and the time of each observation is given by and ordered as  $t_0 < t_1 < t_2 \dots < t_N$ .

The error terms  $\epsilon_j$ , are assumed to be independent, identically distributed (i.i.d) random variables, each with zero mean and finite variance. No further assumptions are made regarding the distribution of  $\epsilon_j$ . In particular, it is not assumed that the distribution is normal. Thus,  $E[\epsilon_j] = 0$  and  $\text{var}(\epsilon_j) = \sigma^2$ . Note that, the i.i.d assumption implies that the variance for each error term  $\epsilon_j$  is identical. We therefore have

$$E[Y_j] = p(t_j, \omega_0) \text{ and } \text{var}(Y_j) = \sigma^2 \text{ for } j = 1, 2, 3, \dots, N$$

For a set of observations  $\{Y_j\}_{j=1}^N$ , produced by the statistical model (2.2), we define the statistical estimator  $\omega_{OLS}$  as

$$\omega_{OLS}(Y) = \arg \min_{\omega \in \Omega} \sum_{j=1}^{j=N} [Y_j - p(t_j, \omega)]^2. \quad (2.4)$$

where  $\Omega \subset \mathbb{R}$  is defined as the physically and biologically feasible region for the parameter  $\omega$ . In other words, the statistical estimator is the solution of the following equations

$$\sum_{j=1}^{j=N} [Y_j - p(t_j, \omega)] \frac{\partial}{\partial \omega} p(t_j, \omega) = 0. \quad (2.5)$$

It is clear that the statistical estimator  $\omega_{OLS}$  is a random variable since each error term  $\epsilon_j$  is a random variable. Furthermore, the estimator attempts to minimize the distance between the observed weekly cumulative number of recovered host individuals and the predictions of model (1.1). A subsequent detailed description of how to obtain the probability distribution associated with  $\omega_{OLS}$  is given in [30]. Our goal in the current study will be to obtain the mean of the probability distribution of  $\omega_{OLS}$ .

#### Uncertainty and sensitivity analysis of $R_0$

As mentioned previously, the basic reproduction number  $R_0$  for the deterministic model (1.1) is given by (1.2) Thus, the value of  $R_0$  depends on the variables  $C_{HV}, \Pi_H, \mu, \sigma_H, \delta_H, \tau_H, \Pi_V, \sigma_V, \mu_V$  and  $\delta_V$ . While deterministic models implicitly assume that the model parameters are not stochastic in nature, an element of uncertainty

is always associated with estimates of these parameters due to factors such as natural variation, errors in measurements and lack of measuring techniques. In general, uncertainty analysis quantifies the degree of confidence in the parameter estimates by producing 95% confidence intervals (CI) which can be interpreted as intervals containing 95% of future estimates when the same assumptions are made and the only noise source is observation error. Additionally, sensitivity analysis identifies critical model parameters and quantifies the impact of each input parameter on the value of an output. In this section, we shall perform uncertainty and sensitivity analysis of the basic reproduction number  $R_0$ . A detailed description of the history and methodology of uncertainty and sensitivity analysis is given in [35].

We will use Latin hypercube sampling (LHS) [35] to quantify the uncertainty in and sensitivity of  $R_0$  as a function of the 7 model parameters ( $C_{HV}, \mu, \sigma_H, \delta_H, \tau_H, \sigma_V, \mu_V$  and  $\delta_V$ ). It is assumed, following [33], that the recruitment rates  $\Pi_V$  and  $\Pi_H$  are constants. This will enable us to obtain CI for the value of  $R_0$  that we have estimated in the last section. For the sensitivity analysis, the partial rank correlation coefficient (PRCC) technique [35] will be used to assess the impact of changes in parameter values on the value of  $R_0$ . PRCC, which uses rank transformation of the data to reduce the effects of non-linearity, provides a measure of monotonicity after the removal of the linear effects of all but one model parameter. PRCC is therefore a global sensitivity analysis technique. The assumed distributions of the model parameters used in the two analyses are mentioned in Table 4.

#### Generalized least squares estimation

The Ordinary Least Squares Estimation (OLS) scheme we employed in the previous section assumed that the variances associated with the epidemic observations were longitudinally constant and not dependent on the values of the observations. This may not be a realistic assumption especially if the epidemic data is influenced by a source

**Table 4 Assumed probability distributions for the parameters of the model (1.1) used in the sensitivity and uncertainty analysis**

Parameter	Mean	Variance	(Dist)
$\mu$	$\frac{1}{3494}$	$\frac{1}{10^5}$	( $\mathcal{N}$ )
$\sigma_H$	1.4	0.35	( $\mathcal{G}$ )
$\delta_H$	0.0035	0.0005	( $\mathcal{N}$ )
$\tau_H$	1	0.2	( $\mathcal{G}$ )
$\sigma_V$	1.4	0.2	( $\mathcal{G}$ )
$\mu_V$	0.5	0.1	( $\mathcal{N}$ )
$\delta_V$	0.01	0.001	( $\mathcal{N}$ )

of non-constant systematic error such as under-reporting. Indeed, under-reporting of dengue cases has been well documented in previous studies such as [36] and [37].

If indeed the epidemic data that is being used in the current study is influenced by under-reporting then the assumption of constant variances associated with the observations is not correct since observation errors will now be proportional to the size of the measurement. Hence, we must use a statistical model, which assumes longitudinally non-constant and model dependent variances for the epidemic observations. In this section therefore, we will attempt to estimate the effective contact rate by fitting the best trajectory of model (1.1) to the epidemic data using a generalized least squares (GLS) scheme. An excellent discussion on the use of the OLS and GLS scheme and different statistical models depending on the assumptions regarding the error present in the observation process is given in [38].

Once again we shall employ the notation and methodology developed in [30]. Apart from the assumptions of the statistical model, as before, we will assume that all reported cases recovered from the infection after a time lag of two weeks and that therefore, the epidemic data, after a lapse of two weeks, represents the total recovered host population. Furthermore, we will again assume that the effective contact rate  $C_{HV}$  is a function of time. Thus, mathematically, Eq. 2.1 where  $t^*$  is the time at which control measures were first implemented. As before, we have assumed that  $t^* = 8$  weeks.

Again, following [30], we will employ inverse problem methodology to obtain estimates of  $\omega = C_{HV}$  by minimizing the difference between the observed weekly cumulative number of recovered host individuals and the model predictions using a generalized least squares (GLS) criterion. This will be done in the context of a statistical model, which assumes that the error in the observation process is directly proportional to the values of the measurement. Hence, it is assumed that

$$Y_j = p(t_j, \omega_0) + p(t_j, \omega_0)\epsilon_j \quad \text{for } j = 1, 2, 3, \dots, N \quad (2.6)$$

where denotes where  $p(t_j, \omega_0)$  denotes the total number of infected individuals who recover in the time span of week  $j$  using the 'true' parameter  $\omega_0$ . Thus, it is assumed that the observed epidemic data is a particular realization of the statistical model (4.1). Under the framework of model (1.1) therefore,

$$p(t_j, \omega_0) = \int_{t_{j-1}}^{t_j} \tau_H I_H(t, \omega_0) \quad \text{for } j = 1, 2, 3, \dots, N \quad (2.7)$$

where  $t_0$  denotes the time of start of the statistical process and the time of each observation is given by and ordered as  $t_0 < t_1 < t_2 \dots < t_N$ .

The rest of the analysis is similar to the method outlined in the previous section and follows easily.

### Methods and materials for statistical inference using the direct-transmission model

#### The direct transmission epidemic model

Several existing studies on the transmission dynamics of dengue use a direct transmission SEIR model [13]. The direct transmission models can be obtained using an approximation to vector-host models such as model (1.1). First, it is assumed that the vector force of infection can be approximated by solving for the equilibrium values of the vector population compartments. Furthermore, it is assumed that the susceptible vector population is approximately a linear multiple of the total host population. These two assumptions effectively result in a rescaling of the host effective contact rate  $C_{HV}$  of model (1.1) into a direct transmission contact parameter  $\beta$ . Using the aforementioned approximation, we formulate a standard incidence, direct transmission SEIR model. More details of the approximation are given in [13].

Mathematically, the direct-transmission model is a follows

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{\beta S(t)I(t)}{N(t)} - \mu S(t) \\ \frac{dE(t)}{dt} &= \frac{\beta S(t)I(t)}{N(t)} - (\sigma + \mu)E(t) \\ \frac{dI(t)}{dt} &= \sigma E(t) - (\tau + \mu + \delta)I(t) \\ \frac{dR(t)}{dt} &= \tau I(t) - \mu R(t) \end{aligned} \quad (3.1)$$

where

$$N(t) = S(t) + E(t) + I(t) + R(t)$$

A description of the variables and parameters of the model (3.1) is given in Table 5 and Table 6 respectively.

For the direct transmission model (3.1), the basic reproduction number  $R_0$  is defined as the total number of secondary infections produced by introducing a single infected host in a completely susceptible population.

**Table 5 Description of the variables of the direct-transmission model (3.1)**

Variable	Description
$N(t)$	Total host population
$S(t)$	Population of susceptible hosts
$E(t)$	Population of exposed hosts
$I(t)$	Population of infected hosts
$R(t)$	Population of recovered hosts

**Table 6 Description of the parameters of the direct-transmission model (3.1)**

Parameter	Description
$\frac{1}{\mu}$	Host mortality
$\delta$	Host disease-induced death rate
$\frac{1}{\sigma}$	Latency period for exposed hosts
$\frac{1}{\tau}$	Recovery time for infected hosts
$\beta$	Effective contact rate for the direct transmission model

Therefore, for model (3.1), the basic reproduction number  $R_0$  is given by:

$$R_0 = \frac{\beta\sigma}{(\tau + \delta + \mu)(\sigma + \mu)}$$

As discussed previously, the existing literature on dengue fever provides excellent estimates of all the parameters of Model (3.1) with the exception of the contact rate  $\beta$ . Hence, our aim will be to estimate the contact rate  $\beta$  using statistical inference and thereby estimate the basic reproduction number  $R_0$ . The basic methodology used for both the OLS and GLS schemes will be similar to the process outlined in the previous sections.

In order to account for the effect of control measures that were implemented during the actual epidemic, we have assumed that the contact rate  $\beta(t)$  is a function of time  $t$ . We will assume that the contact rate was constant up until the point when control measures were implemented, whereupon it changed to a different constant value. Hence, mathematically,

$$\beta(t) = \begin{cases} \beta_1 & t < t^* \\ \beta_2 & t \geq t^* \end{cases} \quad (3.2)$$

where  $t^*$  is the time at which control measures were first implemented. For the purpose of this study and in view of no concrete information being available in this regard, we have assumed that  $t^* = 6$  weeks. We observe that  $\beta(t)$  is most likely not a continuous function of time  $t$ . An alternative definition of the transmission rate  $\beta(t)$ , as a continuous function of time, is given in [33].

### The stochastic model and Markov Chain Monte Carlo (MCMC)

In this section we formulate a stochastic, direct-transmission, discrete-time, (S)usceptible, (E)xposed, (I)nfectious and (R)ecovered/(R)emoved (SEIR) model for the transmission dynamics of dengue virus. We then use standard Markov chain Monte Carlo (MCMC) methods to perform Bayesian Inference on the epidemic data to obtain estimates of the basic reproduction number  $R_0$ . As mentioned in [13], a number of studies exist on the transmission dynamics of dengue that assume direct-transmission. Moreover, a simple approximation can be

used to reduce a vector-host model for dengue virus to a direct transmission model (see [13] and the references given therein for more details). The purpose of using a direct-transmission model is to make stochastic inference computationally tractable. For the purpose of this study, we will broadly follow the procedure outlined in [33].

### Stochastic model formulation

The stochastic SEIR model we will consider is a discrete-time model that has been formulated and discussed previously in [25,33]. Let  $S(t), E(t), I(t)$  and  $R(t)$  denote the susceptible host, exposed host, infected host and removed or recovered host populations at time  $t$  respectively. As is common, our model will assume homogenous mixing of all individuals. Furthermore, consider a time interval  $(t, t + h]$ , where  $h$  denotes the length of time between two observations of the epidemic. In this case therefore,  $h$  is one week. Now, let  $B(t)$  denote the number of susceptible individuals who have contracted disease,  $C(t)$  the number of exposed individuals who have become infected and  $D(t)$  the number of infected individuals who die or have recovered from the disease during this time interval. For the sake of simplicity, and keeping in view the low mortality rate associated with dengue fever, we will assume that the disease-induced death rate is negligible. Following [33], and in view of the fact that the dynamics of dengue disease take place on a much smaller time scale than the average human life expectancy, we will assume that the total population  $N$  remains constant. Thus, mathematically the stochastic model is given by

$$\begin{aligned} S(t+h) &= S(t) - B(t) \\ E(t+h) &= E(t) + B(t) - C(t) \\ I(t+h) &= I(t) + C(t) - D(t) \\ S(t) + E(t) + I(t) + R(t) &= N \end{aligned} \quad (4.1)$$

where,

$$\begin{aligned} B(t) &\sim \text{Bin}(S(t), \lambda(t)), \quad C(t) \sim \text{Bin}(E(t), k_\sigma) \text{ and} \\ B(t) &\sim \text{Bin}(I(t), k_\tau) \end{aligned}$$

are random variables with binomial distributions. The probability of success for these binomial random variables is given by

$$\begin{aligned} \lambda(t) &= 1 - \exp\left(-\frac{\beta(t)I(t)}{N}h\right) \\ k_\sigma &= 1 - \exp(-\sigma h) \\ k_\tau &= 1 - \exp(-\tau h) \end{aligned} \quad (4.2)$$

Here,  $\beta(t), \frac{1}{\sigma}$  and  $\frac{1}{\tau}$  are the time dependent transmission rate, the mean latency period and the mean infectious period respectively. Thus, in model (4.1) the transitions from one compartment to another are formulated as an exponentially distributed stochastic movements. The

probability that each individual will stay in a specific compartment for a time period  $h$  is given by  $\exp(-\Pi h)$ , where  $\Pi$  is the compartment specific movement rate. The binomial distributions in (4.2) are then obtained by summing over the individual Bernoulli trials for every individual in the compartment. It is assumed that each trial is independent and identical for every member of the compartment.

Similar to the previous section, we will assume that the contact rate  $\beta(t)$  is a function of time  $t$ . Thus, mathematically, Eq. (3.2) where  $t^*$  is the time at which control measures were first implemented. As before, we have assumed that  $t^* = 8$  weeks.

The stochastic model (4.1) makes use of the parameters  $\beta(t)$ ,  $\sigma$  and  $\tau$ . Of these parameters, estimates of  $\sigma$  and  $\tau$  can be obtained from existing studies on dengue virus and have been given previously. Therefore, our aim will be to estimate the parameter  $\beta(t)$  using the epidemic data and estimates of initial population sizes. For this purpose, we define the parameter vector, which we will attempt to estimate, as  $\Theta = \beta(t)$ . Moreover, we define

$$\begin{aligned} \mathbf{B} &= \{B(t)\}_{t=0}^{t=\tau^*} \\ \mathbf{C} &= \{C(t)\}_{t=0}^{t=\tau^*} \\ \mathbf{D} &= \{D(t)\}_{t=0}^{t=\tau^*} \end{aligned} \quad (4.3)$$

where  $\tau^*$  denotes the time at which observations of the epidemic have finished.

Thus, based on the available epidemic data, we have complete knowledge of both  $\mathbf{C}$  and  $\mathbf{D}$  but no knowledge of  $\mathbf{B}$ . This lack of knowledge will be a major cause of uncertainty in our analysis. Nevertheless, we will attempt to estimate  $R_0$  for both the time period before control measures are implemented and the time period after control measures are implemented using our knowledge of both  $\mathbf{C}$  and  $\mathbf{D}$ .

### Inference methodology

Based on the definitions given in (4.1) and (4.2), we observe that  $B(t)$ ,  $C(t)$  and  $D(t)$  are conditionally independent random variables. Thus, the likelihood function for the data set  $\{\mathbf{B}, \mathbf{C}, \mathbf{D}\}$  is given by

$$L(\mathbf{B}, \mathbf{C}, \mathbf{D}|\Theta) = \prod_{t=0}^{t=\tau^*} f_1(B(t)|\cdot) f_2(C(t)|\cdot) f_3(D(t)|\cdot) \quad (4.4)$$

where  $f_1, f_2$  and  $f_3$  are the binomial transition probabilities given in (4.1) and (4.2), conditioned on  $\Theta$  and all the epidemic data represented by  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  up until time  $t$ . Therefore, the maximum likelihood estimator for the parameter vector  $\Theta$ , and by extension for  $\beta(t)$  and  $R_0$  can be obtained by maximizing the expression in (4.4).

According to model (4.1), the time series for  $S(t)$ ,  $E(t)$ ,  $I(t)$  and  $R(t)$  can be obtained using  $B(t)$ ,  $C(t)$  and  $D(t)$ . Unfortunately as mentioned previously,  $B(t)$  is unknown since the process of infection is not observed.

Hence, we must also impute the values of  $B(t)$ . These values can then be used to construct the time series for  $S(t)$  and  $E(t)$ .

Since, the likelihood function for  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$ , denoted as  $L(\mathbf{B}, \mathbf{C}, \mathbf{D}|\Theta)$ , is given by (4.4), we can use Bayes' Theorem to obtain, up to a constant, the required posterior distribution that we wish to sample from. This is given by

$$L(\Theta, \mathbf{B}|\mathbf{C}, \mathbf{D}) \propto L(\mathbf{B}, \mathbf{C}, \mathbf{D}|\Theta)\pi(\Theta) \quad (4.5)$$

where  $\pi(\Theta)$  is the prior distribution. Thus, our MCMC algorithm will sample from the conditional probability distributions  $\pi(\Theta|\mathbf{B}, \mathbf{C}, \mathbf{D})$  and  $\pi(\mathbf{B}|\Theta, \mathbf{C}, \mathbf{D})$  to produce samples from the required distribution  $\pi(\Theta, \mathbf{B}|\mathbf{C}, \mathbf{D})$ . In short, our general algorithm will proceed as follows

- Initialize the set  $\mathbf{B}$  using any appropriate initial vector.
- Since,  $\mathbf{C}$  and  $\mathbf{D}$  are known, construct the time series for  $S(t)$ ,  $E(t)$ ,  $I(t)$  and  $R(t)$ .
- Initialize the parameter vector  $\Theta$ .
- Update  $\mathbf{B}$  using the conditional distribution  $\pi(\mathbf{B}|\Theta, \mathbf{C}, \mathbf{D})$ .
- Reconstruct the new time series for  $S(t)$ ,  $E(t)$ ,  $I(t)$  and  $R(t)$ .
- Update  $\Theta$  using the conditional distribution  $\pi(\Theta|\mathbf{B}, \mathbf{C}, \mathbf{D})$ .
- Repeat steps 4 – 6 until the Markov chain has converged and subsequently, the required samples have been obtained.

To sample from  $\pi(\mathbf{B}|\Theta, \mathbf{C}, \mathbf{D})$  one can use the conditional binomial distribution for  $\mathbf{B}$ , making sure that the choice is consistent with the final size and length of the epidemic. This is however computationally very inefficient as most of the draws would be rejected due to the consistency condition. To avoid this issue we condition the proposal on the observed extinction time, following the method described in [33] for computationally efficient sampling.  $\pi(\Theta|\mathbf{B}, \mathbf{C}, \mathbf{D})$  is updated using a random walk proposal.

### Inference from the observed dengue data

An important question that arises at this point pertains to the meaning and significance of the basic reproduction

**Table 7 Posterior mean of the contact rate and basic reproduction number for the Stochastic direct-transmission model (4.1)**

Parameter	Posterior mean
$\beta_1$	3.0650 week <sup>-1</sup>
$\beta_2$	0.6318 week <sup>-1</sup>
$R_0$ before control measures	3.0528
$R_0$ after control measures	0.6293

number  $R_0$  for the stochastic SEIR model. As mentioned previously and as discussed in detail in [31], the basic reproduction number  $R_0$  for the deterministic SEIR model is essentially a threshold quantity which determines the possibility of an outbreak of the disease. Thus, for the deterministic SEIR model, if  $R_0$  is less than unity there is no epidemic while if  $R_0$  is greater than unity there will be a disease epidemic.

Unfortunately, the threshold dynamics of the stochastic SEIR model are not the same. It can be proven that in contrast to the deterministic model, the stochastic SEIR model predicts disease extinction regardless of the value of  $R_0$ . This results in difficulty regarding the interpretation of  $R_0$  as a threshold quantity. Therefore, it is tempting to ask the question: what is the importance of  $R_0$  in the stochastic SEIR model? An answer to this question may be conjectured (but not proven) by referring to [39]. It is proven in [39] that for the stochastic SI model, on average no epidemic will occur if  $R_0 < 1$ , while for  $R_0 > 1$  there is a finite probability that an endemic quasi-equilibrium will develop. We conjecture that this result also holds true for the stochastic SEIR model and that it can therefore be used to explain the significance of  $R_0$  as a threshold quantity for the stochastic SEIR model.

Using the MCMC algorithm discussed in the previous section, we estimate the transmission rate  $\beta(t)$  and hence the basic reproduction number  $R_0$ , both for the time period before control measures are implemented and the time period after the control measures are implemented. We have taken  $t^* = 8$  weeks and the initial population sizes to be the same as in the case of inference from the deterministic model using the GLS and OLS schemes. Furthermore, we have taken a uniform prior distribution for the parameter vector  $\Theta$ . We observe that the results of the MCMC algorithm, displayed in Table 7, are in close agreement with the corresponding results obtained from application of the OLS and GLS schemes to the direct-transmission deterministic model.

## Additional file

Additional file 1: Multilingual abstracts in the six official working languages of the United Nations.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

MI and AK gave initial input and directed the research. MH implemented the algorithms, carried out the programming and drafted the manuscript. AK interpreted the results and edited the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

The authors would like to acknowledge and thank the Punjab Disaster Management Authority for providing the data used for the research work being presented in this article.

## Author details

<sup>1</sup>Department of Mathematics, Lahore University of Management Sciences, DHA, Lahore, Pakistan. <sup>2</sup>Department of Mathematics, Swiss Federal Institute of Technology (ETH), Zurich, Rämistrasse 101, 8092 Zurich, Switzerland.

Received: 26 September 2013 Accepted: 6 March 2014

Published: 7 April 2014

## References

1. Ranjit S, Kisson N: **Dengue hemorrhagic fever and shock syndromes.** *Pediatr Crit Care Med* 2011, **12**:90–100.
2. **World Health Organization: dengue and severe dengue fact sheet.** 2012. [http://www.who.int/mediacentre/factsheets/fs117/en/]
3. Gubler DJ: **Dengue and dengue hemorrhagic fever.** *Clin Microbiol Rev* 1998, **11**(3):480–496.
4. Halstead S, Nimmannitya S, Cohen S: **Observations related to pathogenesis of dengue hemorrhagic fever. IV. Relation of disease severity to antibody response and virus recovered.** *Yale J Biol Med* 1970, **42**(5):311–322.
5. Kautner I, Robinson MJ, Kuhnle U: **Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis, and prevention.** *J Pediatr* 1997, **131**(4):516–524.
6. Shekhar C: **Deadly dengue: new vaccines promise to tackle this escalating global menace.** *Chem Biol* 2007, **14**(8):871–872.
7. Holmes EC, Twiddy SS: **The origin, emergence and evolutionary genetics of dengue virus.** *Infect Genet Evol* 2003, **3**:19–28.
8. Whitehorn J, Farrar J: **Dengue.** *Br Med Bull* 2010, **95**:161–173.
9. Gubler D, Kuno G: *Dengue and Dengue Hemorrhagic Fever.* London: CAB INTERNATIONAL; 1997.
10. Kawaguchi I, Sasaki A, Boots M: **Why are dengue virus serotypes so distantly related? Enhancement and limiting serotype similarity between dengue virus strains.** *Proc R Soc Lond B Biol Sci* 2003, **270**(1530):2241–2247.
11. Garba SM, Gumel AB: **Abu Bakar MR: Backward bifurcations in dengue transmission dynamics.** *Math Biosci* 2008, **215**:11–25.
12. Garba S, Gumel A: **Effect of cross-immunity on the transmission dynamics of two strains of dengue.** *Int J Comput Math* 2010, **87**(10):2361–2384.
13. Wearing HJ, Rohani P: **Ecological and immunological determinants of dengue epidemics.** *Proc Natl Acad Sci* 2006, **103**(31):11802–11807.
14. Esteve L, Vargas C: **Coexistence of different serotypes of dengue virus.** *J Math Biol* 2003, **46**:31–47.
15. Ferguson N, Anderson R, Gupta S: **The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens.** *Proc Natl Acad Sci* 1999, **96**(2):790–794.
16. Esteve L, Vargas C: **A model for dengue disease with variable human population.** *J Math Biol* 1999, **38**(3):220–240.
17. Esteve L, Vargas C: **Analysis of a dengue disease transmission model.** *Math Biosci* 1998, **150**(2):131–151.
18. Chowell G, Diaz-Dueñas P, Miller J, Alcazar-Velazco A, Hyman J, Fenimore P, Castillo-Chavez C: **Estimation of the reproduction number of dengue fever from spatial epidemic data.** *Math Biosci* 2007, **208**(2):571–589.
19. Allen LJ: **An introduction to stochastic epidemic models.** In *Mathematical Epidemiology, Volume 1945 of Lecture Notes in Mathematics.* Edited by Brauer F, Driessche P, Wu J. Springer Berlin Heidelberg, 14197 Berlin Germany; 2008:81–130.
20. Keeling MJ, Ross JV: **On methods for studying stochastic disease dynamics.** *J R Soc Interface* 2008, **5**(19):171–181.
21. Bailey NT: **A simple stochastic epidemic.** *Biometrika* 1950, **37**:193–202.
22. Allen LJ, Flores DA, Ratnayake RK, Herbold JR: **Discrete-time deterministic and stochastic models for the spread of rabies.** *Appl Math Comput* 2002, **132**(2):271–292.
23. Weiss GH, Dishon M: **On the asymptotic behavior of the stochastic and deterministic models of an epidemic.** *Math Biosci* 1971, **11**(3):261–265.
24. Tuite AR, Tien J, Eisenberg M, Earn DJ, Ma J, Fisman DN: **Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions.** *Ann Intern Med* 2011, **154**(9):593–601.
25. Allen L, Driessche P: **Stochastic epidemic models with a backward bifurcation.** *Math Biosci Eng* 2006, **3**(3):445.



26. de Souza DR, Tomé T, Pinho ST, Barreto FR, de Oliveira MJ: **Stochastic dynamics of dengue epidemics.** *Phys Rev E* 2013, **87**:012709.
27. Spencer S: *Stochastic epidemic models for emerging diseases.* PhD thesis: University of Nottingham; 2008.
28. Allen LJ: *An Introduction to Stochastic Processes with Applications to Biology.* New Jersey: Pearson Education; 2003.
29. Allen LJ, Burgin AM: **Comparison of deterministic and stochastic SIS and SIR models in discrete time.** *Math Biosci* 2000, **163**:1–33.
30. Cintrón-Arias A, Castillo-Chávez C, Bettencourt LM, Lloyd AL, Banks H: **The estimation of the effective reproductive number from disease outbreak data.** *Math Biosci Eng* 2009, **6**(2):261–282.
31. Van den Driessche P, Watmough J: **Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission.** *Math Biosci* 2002, **180**:29–48.
32. Chowell G, Hengartner N, Castillo-Chavez C, Fenimore F, Hyman J: **The basic reproductive number of Ebola and effects of public health measures: the cases of Congo and Uganda.** *J Theor Biol* 2004, **229**:119–126.
33. Lekone PE, Finkenstädt BF: **Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study.** *Biometrics* 2006, **62**(4):1170–1177.
34. O'Neill P, Roberts GO: **Bayesian inference for partially observed stochastic epidemics.** *J R Statist Soc A* 1999, **162**:121–129.
35. Sanchez MA, Blower SM: **Uncertainty and sensitivity analysis of the basic reproductive rate: tuberculosis as an example.** *Am J Epidemiol* 1997, **145**(12):1127–1137.
36. Suaya JA, Shepard DS, Beatty ME: **Dengue: burden of disease and costs of illness.** In *TDR. Report of the Scientific Working Group Meeting on Dengue.* Geneva Switzerland: World Health Organization; 2006:35–49.
37. Beatty ME, Beutels P, Meltzer MI, Shepard DS, Hombach J, Hutubessy R, Dessis D, Coudeville L, Dervaux B, Wichmann O, Margolis HS, Kuritsky JN: **Health economics of dengue: a systematic literature review and expert panel's assessment.** *Am J Trop Med Hyg* 2011, **84**(3):473–488.
38. Banks HT, Davidian M, Jr Samuels JR, Sutton KL: *An Inverse Problem Statistical Methodology Summary.* 3994 AK Houten Netherland; 2009.
39. Jacquez JA, O'Neill P: **Reproduction numbers and thresholds in stochastic epidemic models I. Homogeneous populations.** *Math Biosci* 1991, **107**(2):161–186.

doi:10.1186/2049-9957-3-12

Cite this article as: Khan et al.: Estimating the basic reproduction number for single-strain dengue fever epidemics. *Infectious Diseases of Poverty* 2014 **3**:12.

Submit your next manuscript to BioMed Central  
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

