

# Bayesian modelling of a hepatitis A outbreak in men who have sex with men in Sydney, Australia, 1991/1992

## Original Paper

**Cite this article:** Zhang X-S, Charlett A (2019). Bayesian modelling of a hepatitis A outbreak in men who have sex with men in Sydney, Australia, 1991/1992. *Epidemiology and Infection* **147**, e226, 1–7. <https://doi.org/10.1017/S0950268819001109>

Received: 24 December 2018

Revised: 13 May 2019

Accepted: 25 May 2019

### Key words:

Bayesian inference; men who have sex with men; outbreak analysis; reproduction number; transmission dynamics

### Author for correspondence:

Xu-Sheng Zhang,

E-mail: [xu-sheng.zhang@phe.gov.uk](mailto:xu-sheng.zhang@phe.gov.uk)

X.-S. Zhang and A. Charlett

Department of Statistics, Modeling and Economics, National Infection Service, Centre for Infectious Disease Surveillance and Control, Public Health England, London, UK

### Abstract

To control hepatitis A spread by vaccination, accurate estimation of transmissibility is vital. Regan *et al.* (2016) proposed a model of hepatitis A virus (HAV) transmission and used least squares to calibrate model to the 1991/1992 HAV outbreak in men who have sex with men (MSM) in Sydney, Australia. Based on the estimate of  $R_0$ , they obtained the critical immunity of 70% and showed that when the proportion immune <70%, there is a definite chance for outbreaks to take place. The immunity level from previous surveys ranges from 32% to 64% after 1996 while no outbreaks in Australian MSMs have been reported since 1996. Further noticing the ill-distributed parameters, we argue that their estimate of  $R_0$  is not accurate. In this study, we revisited their model by Bayesian inference, which has privilege over least squares. We obtained the appropriate posterior distributions of parameters and the estimate of  $R_0$  ranges from 1.38 to 2.89, indicating a critical immunity of 65%. The reduction in critical immunity and outbreak probabilities predicts the absence of outbreaks in Australian MSMs since 1996. Our study shows the importance of using appropriate methods to provide reliable and accurate estimates of the model parameters especially the transmissibility.

## Introduction

Hepatitis A is caused by hepatitis A virus (HAV) and occurs worldwide. HAV is a ribonucleic acid virus that replicates in the liver and is shed into the stool. Outbreaks of hepatitis A have been mostly linked to the consumption of uncooked contaminated foods and contaminated waters. However, close person-to-person contacts in families, institutions, child care setting and schools are also an important mode of HAV transmission [1, 2]. One special group of people is men who have sex with men (MSM), and numerous outbreaks of hepatitis A have been reported in MSMs. Recent outbreaks of hepatitis A among MSMs in England, France, Germany, Israel and the Netherlands [3] challenge our countermeasures in controlling the transmission of HAV. Hepatitis A is a vaccine preventable disease and the natural countermeasure is hence vaccinating the groups of people in danger. Herd immunity can help control the transmission by imposing the resistance to the spread of a contagious disease within a population that results if a sufficiently high proportion of individuals are immune to the disease, which can be achieved especially through vaccination. The minimum proportion immune of individuals (i.e. the critical immunity threshold) is determined by transmissibility of the contagious agent [4]. The transmissibility is usually characterized by the basic reproductive number  $R_0$ , the mean number of secondary infectious cases generated by a single primary infectious case introduced into a totally susceptible population [5]. To protect the MSM people against HAV by vaccination, we need to evaluate this crucial parameter. Obviously, its precise and reliable estimate is vital for the reasonable design of effective countermeasures such as vaccination.

Regan *et al.*, [2] proposed a compartmental model to estimate  $R_0$  of HAV infection in a MSM population. They applied the model to the 1991/1992 Sydney outbreak of hepatitis A in which large proportions of the infected individuals were MSMs. By applying the least squares algorithm to fit their model outputs to the outbreak data, they obtained an estimate of  $R_0$  ranging from 1.71 to 3.67 and the critical immunity threshold ranging from 42% to 73% in Sydney MSM population. Therefore, as they claimed in their paper, ‘sustained epidemics cannot occur once the proportion immune to HAV is greater than ~70%’. In other words, epidemics could possibly occur if the proportion immune to HAV cannot reach this level.

The examination [6] of the HAV status of all MSMs seen at a large sexual health clinic in inner Sydney between 1996 and 2012 showed that the proportion immune to HAV increased from about 32% in 1996 to 64% in 2012. The similar levels of proportion immune were also reported in the MSM population in Victoria, Australia [7]. In accordance with their results of the event-driven stochastic model for outbreak analysis (Fig. 3a and 3b of [2]), if the proportion immune ranges from 32% to 64%, there is a definite chance for outbreaks of hepatitis A in

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

MSMs to occur with outbreak sizes ranging from a hundred to a thousand. This theoretical indication appeared not to be in agreement with the actual observation: ‘No outbreaks of hepatitis A in Australian MSMs have been reported since 1996’ [2].

We further noted that among ten parameters of Regan *et al.*'s [2] model, (posterior) distributions of eight parameters diverge to their lower and upper bounds set up by fitting constraints (see their Fig. A2 and Table 2). Another important parameter (‘force of infection per sexual contact per day’), though not diverging to lower and upper bounds, appears to have a bimodal distribution. The distribution of the size of the MSM population converged to the lower bound which is clearly below the range of 5000–10 000 of the MSM population size that was estimated from the survey of [8]. In view of such ill-distributed model parameters, the accuracy and reliability of their estimate of  $R_0$  are doubtful.

We think the abnormal behaviours of the distributions of model parameters obtained in [2] stem from two aspects. The first is that the least squares algorithm is inappropriate for calibration of the transmission model to obtain the reliable estimates of model parameters (see Appendix A); the second is that the only data used in [2] (i.e. the time series of case onset dates) lack enough information to distinguish ten model parameters because of the inherent correlations between them such as transmissibility and susceptibility.

In view of these questions, in this study we revisited the model of [2] by proposing the Bayesian inference for model calibration in replace for the least squares algorithm and reducing the number of independent model parameters by considering the early transmission dynamics of the outbreak. These changes will hopefully improve the behaviour of distributions of model parameters and compromise the discrepancy between the field observations and theoretical prediction from the analysis of [2].

## Models and methods

We adopted the deterministic model of [2] for the transmission dynamics of the 1991–1992 hepatitis A outbreak in Sydney, Australia. Susceptible MSMs ( $S$ ) contact with infections and then become infected but not yet infectious ( $E$ ). The exposed people progress to become occult infectious ( $I_O$ ) at rate  $\sigma$ . A proportion of infections ( $p_S$ ) becomes symptomatic ( $I_S$ ) and the rest still remains asymptomatic but infectious ( $I_A$ ) at rate  $\gamma_1$ . All infections recover and become immune to HAV ( $R$ ) at rate  $\gamma_2$ . The model equations are given in Appendix B.

We propose the Bayesian inference for the model calibration and use Markov Chain Monte-Carlo (MCMC) [9, 10] to sample posterior distributions of model parameters and hence to obtain their median and 95% credible interval (CI). We think Bayesian inference has privilege over least squares algorithm in calibrating the transmission model and obtaining the estimates of model parameters. In order to consider the variation in the number of monthly cases, Regan *et al.*, [2] added Poisson noise to the observed epidemic data. Within the Bayesian framework through MCMC, the variation in the data has been naturally included in the likelihood function (e.g., negative binomial distribution we use in this study). Moreover, Bayesian inference is more flexible and realizes more information by sampling the posterior distributions of model parameters (the detail of the Bayesian inference method is given in Appendix B).

We can also reduce the number of independent model parameters by considering the early stage of transmission dynamics. Following [11], we assume that at the early stage of the outbreak, the number

of newly infected cases increases exponentially with growth rate  $\Psi_r$ . Then the initial conditions can be parameterized as follows:

$$E(0) = \frac{(\alpha_0 + (\alpha + \gamma_1)\delta t)}{\sigma\delta t} I_O(0),$$

$$I_S(0) = \frac{p_S\gamma_1\delta t}{(\alpha_0 + (\alpha + \gamma_2)\delta t)} I_O(0),$$

$$I_A(0) = \frac{(1 - p_S)\gamma_1\delta t}{(\alpha_0 + (\alpha + \gamma_2)\delta t)} I_O(0)$$

and  $S(0) = N\rho - E(0) - I_O(0) - I_S(0) - I_A(0)$

Here  $\alpha_0 = \exp(\Psi_r\delta t) - 1$  and  $\delta t$  is a small time step used in the simulations. We assume the transmission dynamics started from December 1990 and  $I_O(0)$  represents the number of occult infections on December 1990.  $N$  is the effective size of MSM population and  $\rho$  is the initial proportion of MSMs that are susceptible to HAV.  $\alpha$  is the rate of entry/exit from MSM population, which is fixed at  $1/(30 \times 365)$  as in [2] who assumed a sexually active period from age 20 to 50 years old.

Following [12], we can establish the relationship between the transmission rate  $\beta$  (i.e. what Regan *et al.*, [2] called ‘force of infection’) and the initial growth rate  $\Psi_r$ ,

$$\beta = \left(1 + \frac{\alpha + \psi_r}{\sigma}\right) \frac{\gamma_1 + \alpha + \psi_r}{1 + \frac{(1 - p_S)\gamma_1}{\gamma_2 + \alpha + \psi_r}}$$

The derivations of these expressions are given in Appendix B. After these treatments, the model is now described by eight-independent parameters (see Table 1).

Even with the above parametrization of the model, there are still a lot of model parameters to be estimated. Because of the inherent associations between model parameters such as the proportion of non-immune people and transmissibility, the calibrating transmission model to symptom onset dates cannot identify all the eight model parameters. In general, the lack of identifiability was met when the calibrating transmission model with one dataset in other infectious diseases (e.g. [13]). To avoid this identifiability problem, we first consider a situation by fixing some model parameters based on the literature.

The survey results of [6] suggested that about 30% of the MSM population in inner Sydney between 1996 and 2012 would have been immune. This was further supported by another survey: Heywood *et al.*, [14] show the seropositivity is 34% in the Australia population in 1988. In view of these, we can fix the initial proportion of the susceptible individuals before the 1991/1992 outbreak at  $\rho = 70\%$ . The risk that a hepatitis A infection will cause symptomatic case is known to increase with age at infection. In accordance with the modelling of [15], the proportion of symptomatic for individuals of age  $>20$  years can be approximated at 85%. Hence we can use this estimate to fix the proportion of symptomatic cases at  $p_S = 85\%$ . To further reduce the number of model parameters to be estimated, we also fix the life history parameter at their middle values of the ranges used in [2]: latent period ( $P_L$ ) =  $1/\sigma = 14$  days, period of occult infection ( $P_{OI}$ ) =  $1/\gamma_1 = 14$  days and period of symptomatic infection ( $P_{SI}$ ) =  $1/\gamma_2 = 7$  days (also see [16]). We now have three parameters to be estimated: size of the MSM population ( $N$ ), the number of initial occult infection ( $I_O$ )

**Table 1.** Prior and posterior distributions of model parameters under different model variants. Here  $U$  stands for the uniform distribution

Parameters	Priors	Posterior			
		Model variant 0	Model variant I	Model variant II	Model variant III
$N$	$U[500, 10\,000]$	1033 [848, 1381]	1299 [824, 1969]	1321 [708, 2280]	1274 [676, 2277]
$P_L$	$U[10, 18]$	14 (fixed)	14 (fixed)	14 (fixed)	13.9 [10.1, 17.9]
$P_{OI}$	$U[10, 18]$	14 (fixed)	14 (fixed)	14 (fixed)	14.6 [10.2, 17.9]
$P_{SI}$	$U[3, 11]$	7 (fixed)	7 (fixed)	7 (fixed)	6.9 [3.2, 10.9]
$p_S$	$U[0.5, 1.0]$	0.85 (fixed)	0.67 [0.51, 0.98]	0.67 [0.50, 0.99]	0.68 [0.51, 0.98]
$\rho$	$U[0.5, 1.0]$	0.70 (fixed)	0.70 (fixed)	0.67 [0.51, 0.98]	0.70 [0.50, 0.99]
$I_0(0)$	$U[0.1, 20]$	2.14 [1.27, 4.03]	2.68 [1.34, 5.38]	2.67 [1.30, 5.61]	2.72 [1.20, 6.07]
$\Psi_r$	$U[0.01, 0.10]$	0.029 [0.025, 0.032]	0.029 [0.025, 0.032]	0.031 [0.013, 0.047]	0.029 [0.012, 0.049]
$\beta$	–	0.13 [0.12, 0.14]	0.12 [0.11, 0.14]	0.13 [0.085, 0.18]	0.125 [0.077, 0.21]
$R_0$	–	2.00 [1.84, 2.12]	2.01 [1.85, 2.14]	2.09 [1.40, 2.29]	2.02 [1.38, 2.89]
$R_e$	–	1.40 [1.29, 1.49]	1.41 [1.30, 1.50]	1.40 [1.29, 1.50]	1.41 [1.27, 1.57]
$\eta$	$U[1.01, 40]$	2.20 [1.22, 4.90]	2.22 [1.26, 5.00]	2.27 [1.23, 4.99]	2.23 [1.22, 4.95]

on Dec 1990 and the initial growth rate  $\Psi_r$ . Uniform distributions within some ranges will be used for priors of model parameters (Table 1). In view of the downsized distribution of the population size of [2], we assume a uniform prior from 500 to 10 000 for  $N$ . The prior for growth rate  $\Psi_r$  is chosen so that a large range of  $R_0$  for HAV will be covered. The priors for other parameters follow that of [2]. The arrangement of the model parameters will be hereafter referred to as model variant 0.

### Sensitivity analysis

To assess how the estimate of  $R_0$  depends on the different choices of life history parameters, the proportion of the susceptible and the proportion of symptomatic cases, we relax the above assumptions of the model parameters and consider the following model variants:

*Variant I:* Allowing the proportion of symptomatic cases ( $p_S$ ) to vary within the range (0.5,1.0) as suggested in [2]. We now have four parameters to be estimated:  $N$ ,  $I_0$ ,  $\Psi_r$  and  $p_S$ .

*Variant II:* Allowing both  $p_S$  and  $\rho$  to vary within the range (0.5,1.0) and fixing life history parameters ( $P_L$ ,  $P_{OI}$  and  $P_{SI}$ ) at 14:14:7.

*Variant III:* The model variant allows eight parameters ( $N$ ,  $P_L$ ,  $P_{OI}$ ,  $P_{SI}$ ,  $\rho$ ,  $p_S$ ,  $I_0(0)$  and  $\Psi_r$ ) to be estimated. We regard the estimates of model variant III as our estimates and use further for our outbreak analysis.

### Outbreak analysis

To capture the stochastic features of outbreak of hepatitis A within a MSM population, we approximate the deterministic equations by the Monte Carlo algorithm [17], which tracks the succession of discrete events that change the number of individuals in each of the six infection compartments. The whole stochastic system is described by 13 transmission and transition events. Each event occurs at a rate equal to that in the deterministic model (see Table A1). The size of each compartment is the number of individuals occupying the compartment. From initial sizes of compartments, the programme first determines the time of the next event, which follows an exponential distribution with mean  $1/\Omega$ ;

here  $\Omega$  denotes the sum of all individual event rates. The nature of the next event is chosen at random, with each of the 13 events having a probability equal to its own rate divided by  $\Omega$ . After each occurrence, the sizes of the compartments are updated according to the picked events (Table A1). The input parameters for these simulations were the 5000 samples of model parameters ( $N$ ,  $P_L$ ,  $P_{OI}$ ,  $P_{SI}$ ,  $p_S$ ,  $I_0(0)$  and  $\Psi_r$ ) obtained from the above model calibration. For each set of model parameters, 20 stochastic realisations are generated to evaluate probabilities and size of outbreaks. To investigate how the proportion immune ( $\rho$ ) impacts the outcome of simulating transmission dynamics, we replace the proportion immune in the input parameters by 0%, 5%, 10%, ..., 100%. We define sustained epidemics as simulations in which the epidemic ceases via depletion of susceptible individuals rather than by chance elimination. Technically, we regard an outbreak takes place when the total number of symptomatic cases exceed 5% of susceptible individuals.

## Results

### Estimation of $R_0$

The estimates of model parameters under four model variants are shown in Table 1. The initial number of occult infections ( $I_0(0)$ ) and the initial growth rate ( $\Psi_r$ ) are nearly insensitive to the model variants, and have medians of 2.1–2.7 and 0.29–0.31, respectively. Under the model variant 0, the three model parameters converge well (see Fig. A1) and the basic reproduction number ( $R_0$ ) is estimated to have median 2.00 and 95% CI from 1.84 to 2.12. Under model variant I where the proportion of symptomatic cases ( $p_S$ ) was also to be estimated, albeit the three parameters (i.e.  $N$ ,  $I_0$  and  $\Psi_r$ ) converging well as in model variant 0, the posterior distribution of parameter  $p_S$  is nearly equivalent to its prior (Fig. A3). To test how the results depend on the choices of the values of life history parameters ( $P_L$ ,  $P_{OI}$  and  $P_{SI}$ ) and the proportion of susceptible ( $\rho$ ), 10 combinations of three life history parameters were sampled by Latin hypercube sampling (see Table 2a). The estimate of  $R_0$  is nearly insensitive to the variation in life history parameters given the proportion of the

**Table 2.** Sensitivity analyses with (a) different combinations of three life history parameters by Latin hypercube sampling and (b) different proportion of the susceptible under model variant I

(a). Proportion of the susceptible ( $\rho$ ) is fixed at 70%							
$L_P:L_{OI}:L_{SI}$	$R_0$	$R_e$	$N$	$p_S$	$\beta$	$I_0(0)$	$\Psi_r$
priors	–	–	[500, 10 000]	[0.5, 1.0]	–	[0.1, 20]	[0.01, 0.10]
14:14:7	2.01 [1.85, 2.14]	1.41 [1.30, 1.50]	1299 [824, 1969]	0.67 [0.51, 0.98]	0.12 [0.11, 0.14]	2.68 [1.34, 5.38]	0.029 [0.025, 0.032]
17:10:6	1.98 [1.83, 2.11]	1.39 [1.28, 1.48]	1305 [856, 1999]	0.68 [0.50, 0.99]	0.17 [0.15, 0.20]	1.89 [0.97, 3.88]	0.030 [0.026, 0.033]
11:15:9	1.97 [1.83, 2.1]	1.39 [1.28, 1.47]	1343 [867, 1982]	0.68 [0.51, 0.99]	0.11 [0.10, 0.13]	2.95 [1.51, 5.89]	0.030 [0.026, 0.033]
10:17:5	1.97 [1.83, 2.1]	1.38 [1.28, 1.47]	1356 [867, 2068]	0.68 [0.51, 0.98]	0.11 [0.10, 0.12]	3.31 [1.68, 6.54]	0.030 [0.026, 0.033]
18:15:7	2.13 [1.95, 2.28]	1.49 [1.37, 1.60]	1149 [760, 1738]	0.69 [0.51, 0.98]	0.13 [0.11, 0.14]	2.68 [1.4, 5.51]	0.027 [0.023, 0.030]
15:11:10	1.99 [1.84, 2.12]	1.40 [1.29, 1.48]	1310 [871, 1923]	0.69 [0.51, 0.99]	0.14 [0.12, 0.18]	2.14 [1.07, 4.35]	0.029 [0.025, 0.033]
13:12:5	1.93 [1.78, 2.05]	1.35 [1.25, 1.43]	1433 [919, 2171]	0.67 [0.51, 0.98]	0.14 [0.13, 0.16]	2.31 [1.14, 4.78]	0.030 [0.026, 0.034]
12:16:4	1.99 [1.85, 2.12]	1.39 [1.29, 1.48]	1330 [847, 1953]	0.67 [0.51, 0.98]	0.12 [0.11, 0.13]	3.14 [1.53, 6.17]	0.029 [0.025, 0.032]
15:13:4	1.99 [1.83, 2.12]	1.39 [1.28, 1.49]	1325 [850, 1938]	0.67 [0.51, 0.98]	0.14 [0.13, 0.16]	2.52 [1.27, 5.23]	0.029 [0.025, 0.032]
13:17:8	2.06 [1.87, 2.2]	1.44 [1.31, 1.54]	1244 [791, 1837]	0.66 [0.50, 0.99]	0.11 [0.09, 0.12]	3.23 [1.61, 6.77]	0.028 [0.024, 0.031]
16:13:10	2.06 [1.89, 2.2]	1.44 [1.32, 1.54]	1258 [801, 1857]	0.66 [0.50, 0.98]	0.13 [0.11, 0.16]	2.52 [1.16, 5.44]	0.028 [0.024, 0.031]
(b). The life history parameters fixed as $P_L = 14$ days, $P_{OI} = 14$ days and $P_{SI} = 7$ days.							
$\rho$	$R_0$	$R_e$	$N$	$p_S$	$\beta$	$I_0(0)$	$\Psi_r$
priors	–	–	[500, 10 000]	[0.5, 1.0]	–	[0.1, 20]	[0.01, 0.10]
50%	2.80 [2.60, 3.0]	1.40 [1.30, 1.50]	1732 [1123, 2603]	0.67 [0.51, 0.99]	0.17 [0.16, 0.20]	2.42 [1.22, 4.87]	0.046 [0.042, 0.051]
60%	2.35 [2.17, 2.5]	1.41 [1.3, 1.5]	1512 [966, 2196]	0.66 [0.51, 0.98]	0.14 [0.13, 0.17]	2.6 [1.29, 5, 5.16]	0.037 [0.033, 0.040]
70%	2.01 [1.85, 2.14]	1.41 [1.30, 1.50]	1299 [824, 1969]	0.67 [0.51, 0.98]	0.12 [0.11, 0.14]	2.68 [1.34, 5.38]	0.029 [0.025, 0.032]
80%	1.75 [1.61, 1.87]	1.40 [1.29, 1.49]	1154 [751, 1710]	0.68 [0.51, 0.98]	0.11 [0.10, 0.13]	2.82 [1.44, 5.79]	0.022 [0.019, 0.025]
90%	1.56 [1.44, 1.67]	1.40 [1.30, 1.51]	1022 [658, 1569]	0.68 [0.51, 0.99]	0.10 [0.09, 0.11]	2.87 [1.4, 5.73]	0.017 [0.014, 0.020]

susceptible ( $\rho$ ) being fixed. Further supplement was provided by choosing  $\rho$  at 50%, 60%, 70%, 80% and 90%. The estimate of  $R_0$  is fairly sensitive to the value of  $\rho$  (see Table 2b). Therefore it is no wonder in Model variant II where parameter  $\rho$  was also to be estimated that the effective reproduction number  $R_e$  converge well even both  $R_0$  and  $\rho$  don't (Fig. A4).

Under the general model choice (Model variant III) the median of  $R_0$  is 2.02 and its 95% CI ranges from 1.38 to 2.89. This very well agrees with the other three simplified model variants. However, Figure 1 shows that within model variant III, the posterior distributions of three life history parameters ( $P_L$ ,  $P_{OI}$ ,  $P_{SI}$ ), and both the proportion of symptomatic cases ( $p_S$ ) and the proportion of the susceptible ( $\rho$ ) are nearly equal to their priors. This result has been demonstrated by the sensitivity analysis shown in Table 2. Table 1 further shows that the distribution of  $R_e$  remains nearly the same over the four model variants. Notice that  $R_e$  is the product of  $R_0$  and the proportion of the susceptible ( $\rho$ ), insensitiveness of  $R_e$  to  $\rho$  indicates the highly negative correlation between  $R_0$  and  $\rho$ . This observation reflects the fact that data of symptom onset dates do not contain sufficient information to distinguish these characteristics. Nevertheless, these posteriors appear to be much better than that obtained by least squares (see Fig. A2 of [2]). Although the posterior distribution of the growth rate does not converge to a bell-shaped curve, it does converge to a restrictive range, and its associated parameter transmission rate ( $\beta$ ) does converge well. The distribution of  $R_0$

was also restricted in a range, however, the distribution of effective reproductive number  $R_e$  converges very well, with a median of 1.44 and 95% CI from 1.27 to 1.57.

The effective size of MSM population in Sydney is estimated to be 1274 with 95% CI from 676 to 2277. This is comparable with that obtained in [2]. The dispersion parameter in the negative binomial likelihood that captures the variation in observed epidemic data (Fig. 2) consistently has a median of 2.2, which show that the monthly counts of symptom onset dates during the outbreak are over-dispersed when compared to a Poisson noise assumed in [2].

The model was also fitted to the total outbreak ( $n = 570$ ) which included a further 240 cases that were recorded as heterosexual men/women and children or unknown. The nearly same estimate of  $R_0$  was obtained (see Table A2). The population size is estimated to be nearly double that for the analysis based on the outbreak of 330 cases.

### Outbreak probability

Based on the estimate of transmissibility within model variant III, the critical level of immune proportion to HAV in the MSM population in Sydney is 65%. The outbreak probability we obtain here is greatly decreased compared to that of Regan *et al.*'s analysis (see Fig. 3a). The outbreak analysis shows that an outbreak probability per initial MSM case is about 1.8% (vs. 12% of [2]) when the population immunity was 65%; it is only 6.6% (vs. 28% of [2]) when the

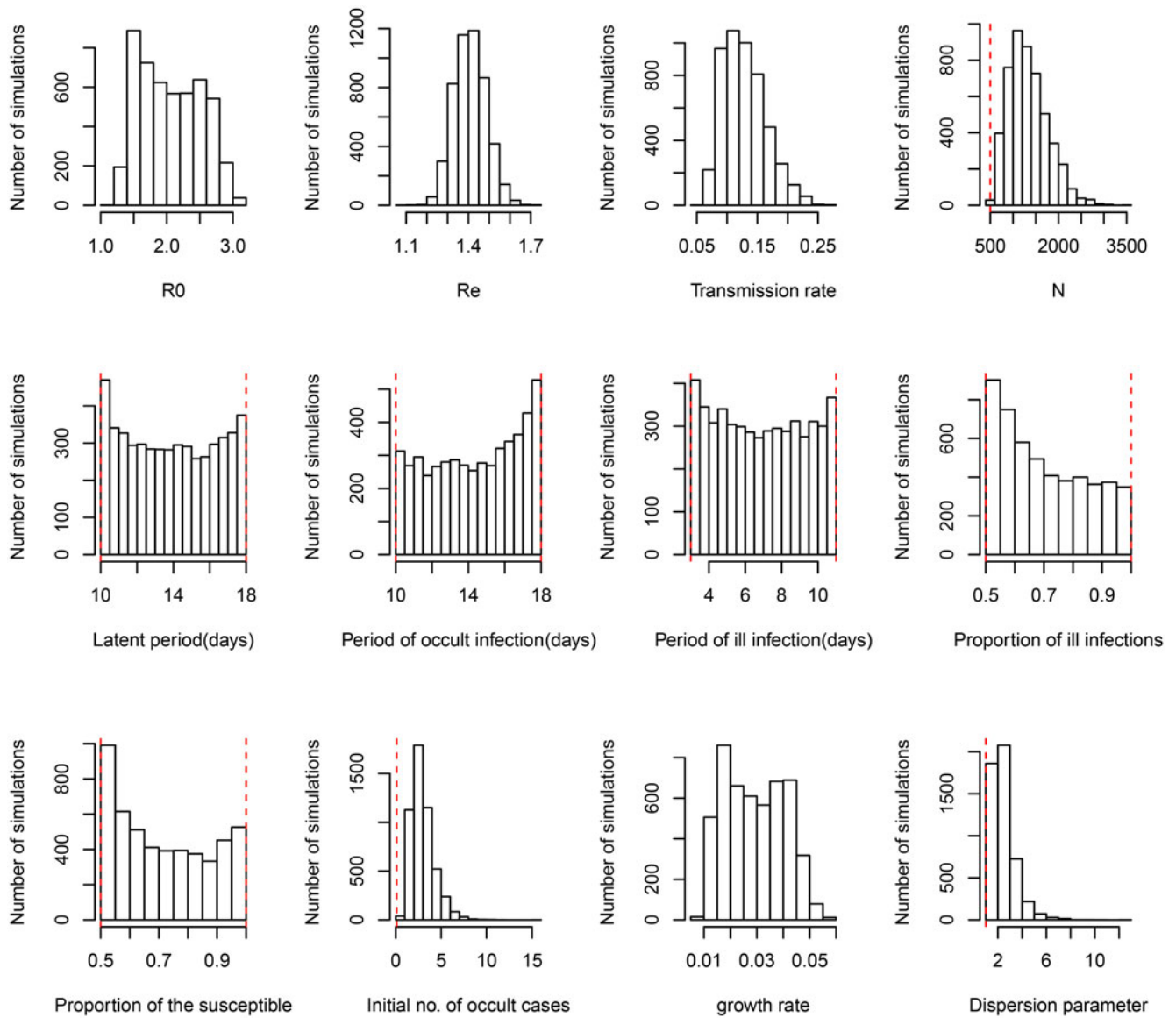


Fig. 1. Posterior distributions of model parameters under model variant III. The red vertical lines represent the lower and upper bounds of uniform priors.

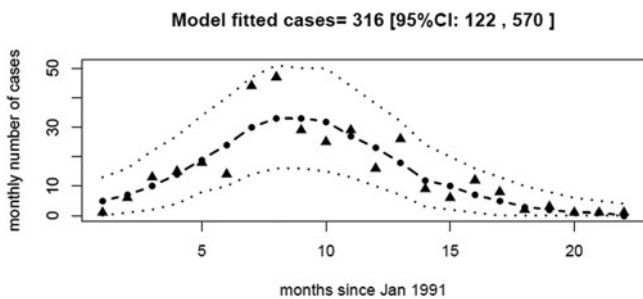


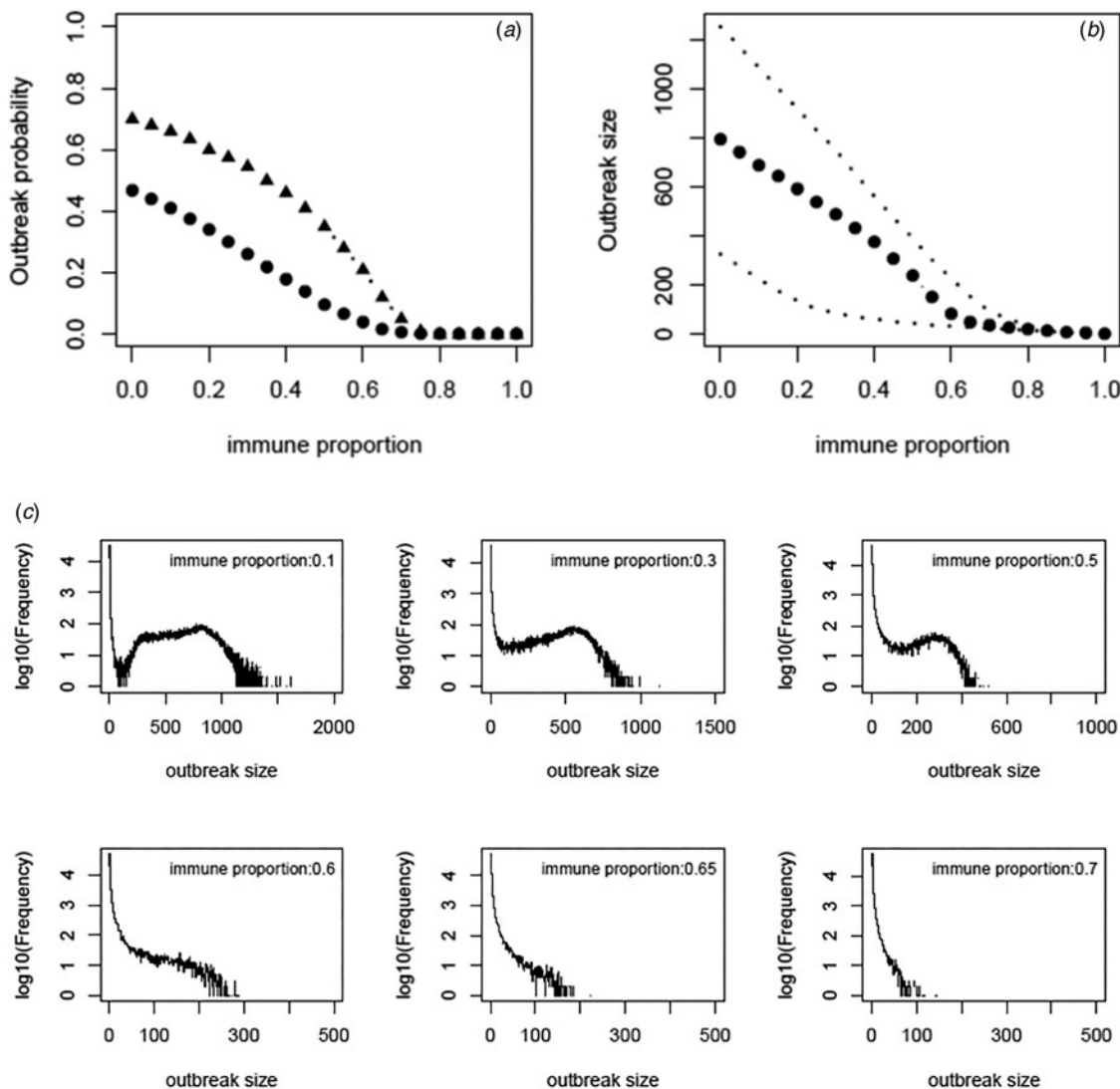
Fig. 2. Model fitting to observed symptom onset dates of 330 MSM patients under model variant III. Triangles represent observed data, dark dashed line represents the median of model predictions and thin dashed lines represent the 95% CIs.

population immunity was 55%. The corresponding sizes of the outbreaks are also reduced substantially (see Fig. 3b). Using the estimates of model parameters based on the total outbreak ( $n = 570$ ),

the outbreak analysis suggested the similar outbreak probabilities to that based on the outbreak of 330 cases (Fig. A7).

### Discussion

We revisited a HAV transmission model of [2] by using Bayesian inference and considering the early stage dynamics of the outbreak to reduce the number of independent model parameters. We, therefore, improve the model analysis and prediction. The posterior distributions of model parameters either converge well to a bell-shaped curve or restrict in a limited range; they are much more proper than that of [2]. The estimate of  $R_0$  ranges from 1.38 to 2.89 and the corresponding critical level of immune proportion to HAV reduce to 65% for controlling the hepatitis A outbreaks in MSMs in Sydney. Although this appears a small reduction in the critical level compared to 70% suggested in [2], the outbreak analysis suggests that the outbreak probabilities under a proportion immune smaller than this critical level decrease greatly (Fig. 3a). A



**Fig. 3.** (a) Outbreak probability as a function of the immune proportion, (b) the median size and 95% CI of such outbreaks and (c) examples of distribution of the outbreak size as a function of immune proportion. In panel (a) triangles represent the outbreak probability predicted by Regan *et al.*, [2].

combination of both reductions greatly compromises the discrepancy between the theoretical prediction of [2] and the observation of no outbreaks in Australian MSMs since 1996.

One concern is on the size of MSM population. Based on the survey of [8], the estimated size of the MSM population in the postcodes that predominated in the 1991 outbreak ranged from 5000 to 10 000. If we fix the population size at such levels and assume the proportion of the susceptible ranging from 0.5 to 1.0, then the simulated transmission dynamics cannot reach the peak around the end of 1991 and downturn will take much longer to emerge than the duration of the outbreak. This is because many susceptible MSMs are still there to be infected. In this simple model proposed by [2], the downforce for the spread of HAV infection is the deletion of susceptible MSMs. By allowing the population size to vary and thus to be estimated, our analysis suggests a size of 1274, which is much smaller than the survey, though being comparable to that of [2]. Another possible downturn force for an outbreak is the reduction in contact rate between MSM people, presumably due to interventions and health education by local health agency since the onset of the outbreak (c.f.,

[16]). A simple simulation analysis (see ‘Effect of behaviour response during the outbreak’ in Technical appendix B) shows that when the contact rate and thus the transmission rate reduce 29% (ranging from 2% to 47%) from July 1991, the estimated size of MSM population increases to 2811, which is still much smaller than the ranges suggested by the survey of [8]. Nevertheless, the shrinkage of the size of the MSM population is another cause that reduces the outbreak probabilities.


This deviation of theoretical estimation from the actual survey in the effective size of MSM population might reflect the fact that the real mixing patterns among MSM population depart significantly from the homogeneous-mixing. In more or less realistic contacts, most people have only one or few contacts while only a very small proportion of people have large numbers of contacts [18]. This deviation might also suggest that there would be a small but highly connected core group of MSMs during the outbreak while the remainder large proportion of MSMs are isolated in some way from the spread of HAV. Such structured contact patterns should greatly reduce the effective size of MSMs in view of spread of HAV infection. Transmission

models that include these realistic demographic characteristics of MSMs and their contact rates will be required for a complete understanding of the spread of HAV especially the force for the downturn of the outbreak within the population where susceptible has not been exhausted.

## Conclusion

In this study, we have demonstrated the importance of using appropriate inference methods for model parametrization and calibration. With the information of only symptom onset dates of symptomatic infections, many epidemiological parameters are left unidentifiable. Making reasonable assumptions about the transmission dynamics can help reduce the number of independent parameters and therefore the uncertainty in model calibration. Bayesian inference through MCMC sampling treats model fitting processes as a dynamic process and can dig out the possible combinations of model parameters that can generate the time series comparable to the observed outbreak dataset. We show that compared to the traditional least squares algorithm, Bayesian inference provides a more reliable estimation of model parameters; the outbreak analysis based on the calibrated model might be more robust to explain the field observation (i.e., the absence of the HAV outbreak in MSM population in Australia since 1996).

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268819001109>

**Author ORCIDs.**  X.-S. Zhang, 0000-0002-6596-4500

**Acknowledgements.** We would like to thank the anonymous reviewer for his/her helpful comments and suggestions. The authors received no specific funding for this work.

**Declaration of interest.** None

## References

1. Wu JB *et al.* (2014) Source identification through social network in an epidemiological investigation of a hepatitis A outbreak at an elementary school in Anhui province, China. *Epidemiology and Infection* **142**, 1450–1458.
2. Regan DG *et al.* (2016) Estimating the critical immunity threshold for preventing hepatitis A outbreaks in men who have sex with men. *Epidemiology and Infection* **144**, 1528–1537.
3. **Special edition:** Outbreak of hepatitis A in men who have sex with men in 2017 December 2017 Eurosurveillance.
4. Vynnycky E and White RG (2010) *An Introduction to Infectious Disease Modelling*. New York: Oxford University Press.
5. Anderson RM and May RM (1991) *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press.
6. Ali H *et al.* (2015) Increasing hepatitis A immunity in men who have sex with men in Sydney, 1996–2012. *Vaccine* **33**, 4745–4747.
7. Weerakoon AP *et al.* (2012) Immunity to hepatitis A when outbreaks of infection in men who have sex with men (MSM) are rare. *Vaccine* **30**, 3430–3434.
8. Madeddu D *et al.* (2006) Estimating population distribution and HIV prevalence among homosexual and bisexual men. *Sexual Health* **3**, 37–43.
9. Metropolis N *et al.* (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.
10. Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
11. Birrell PJ *et al.* (2011) Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18238–18243.
12. Wearing HJ, Rohani P and Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLOS Medicine* **2**, e320.
13. Corbella A *et al.* (2018) Exploiting routinely collected severe case data to monitor and predict influenza outbreaks. *BMC Public Health* **18**, 790.
14. Heywood AE *et al.* (2012) Changes in seroprevalence to hepatitis A in Victoria, Australia: a comparison of three time points. *Vaccine* **30**, 6020–6026.
15. Armstrong GL and Bell BP (2002) Hepatitis A virus infections in the United States: model-based estimates and implications for childhood immunization. *Pediatrics* **109**, 839–845.
16. Zhang X-S and Iacono GL (2018) Estimating human-to-human transmissibility of hepatitis A virus in an outbreak at an elementary school in China. *PLoS ONE* **13**, e0204201.
17. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* **81**, 2340–2361.
18. Bansal S, Grenfell B and Meyers LA (2007) When individual behaviours matters: homogeneous and network models in epidemiology. *Journal of Royal Society Interface* **4**, 879–891.