**Technical Appendix B**

**Model equations of HAV transmission dynamics**

The model is described by the following system of ordinary differential equations

(1)

Where *S* denotes susceptible to infection, *E* exposed to infection (infected) but not yet infectious, *I*O infectious but pre-symptomatic (occult), *I*S symptomatically infected, *I*A asymptomatically infected, and *R* recovered with lifetime immunity to reinfection. As assumed in Regan et al (2016), the severity of symptoms is such that sexual activity is highly unlikely, and the recovery rates for asymptomatic and symptomatic infection are assumed to be equal.

*N* represents the effective size of MSM population involved with 1991/1992 Sydney outbreak of hepatitis. Parameter *α* denotes the rate of entry and exit from MSMS population, which was fixed as 1/(30\*365). That is the sexual active period is 30 years (from 20 to 50) as suggested by Regan et al (2016). Parameter *σ* stands for the progression rate from exposure to occult infectiousness (i.e., PL =1/*σ* is the mean latent period). Regan et al. (2016) assumed that PL ranges from 10 days to 18 days. γ1 stands for progression rate from occult infectiousness to symptomatic infectiousness (i.e. *P*OI =1/γ1 is the period patients staying in occult infection). Regan et al. (2016) assumed that *P*OI ranges from 10 days to 18 days. γ2 stands for progression rate from symptomatic infection to recovery (i.e. *P*SI =1/γ2 is the period patients staying in symptomatic infection). Regan et al. (2016) assumed that *P*SI ranges from 3 days to 11 days. Parameter *p*S stands for the proportion of cases that are symptomatic, and *ρ* stands for the proportion of the susceptible at the beginning of the outbreak.

*Parametrization of the initial condition*: Assuming that the initial infections increase exponentially at the growth rate *Ψ*r :. From this assumption, . Integrating the 3rd equation of system (1) leads to, . Hence we have

.

Here . Similarly we have,

, (2),

and therefore .

The initial condition is set up by three parameters: the initial exponential growth rate (*Ψ*r), the initial number of occult infections (*I*O(0)) and the initial proportion of the susceptible (*ρ*).

The steady-state solution of the system of ODEs (1) can be easily obtained. Here we only list the expression for S\* (the size of the population susceptible to infection at equilibrium):

.

From this, we can obtain the expression of basic reproduction number (the average number of secondary infections that will be occur on average when an infectious individual is introduced into a wholly susceptible population)

(3)

(Vynnycky and White 2010)

Under the special situation where α=0, .

*Relationship between transmission rate and growth rate*: Using the method of Wearing et al (2005), we can establish the relationship between transmission rate *β* and the initial growth rate*Ψ*r. The equations that govern the rate of change in the numbers within infectious state that contribute to the transmission can be written as

From which the characteristics equation at time *t*=0 is,

Its dominant eigenvalue gives

Putting this into equation (3) lead to

Under the special situation where α=0, .

**Bayesian modelling**

As shown in Figure A2 of Regan et al (2016), most of distributions of model parameters do not converge (except for the population size). Their results that were based on these 95% plausible interval are problematic. That is, mostly, the ranges of model parameters were set subjectively and model fitting by minimizing the square distance between data and model outputs did not provide useful information for our understanding of the underlying dynamics of transmission. In this study we suggest using the Bayesian MCMC framework to calibrate the model and thus estimate the relevant parameters.

*Likelihood function*: To reflect the huge dispersion in the weekly number of cases, the negative binomial likelihood function will be assumed to capture the variation. The number *x* (*t*) of cases on week *t* is distributed as

 (4a)

where

 (4b)

Here  is the dispersion parameter and  are the predictions of the cases on week *t* from the transmission dynamics model (1),

(5)

Assuming that the observed weekly incidence *x*(1), *x*(2),…, *x*(*T*) are conditionally independent, the total likelihood given model parameters **Θ** ={*N*,*β*, σ,*γ*1,*γ*2, *p*S,*ρ* ,*Ψ*r,*IO*} is

 (6)

*Inference model*: The prior distributions *f*(**Θ**) of model parameters are extracted from Regan et al (2016) or literature (see Table 1). Employing Bayesian framework through the combination of the prior distribution *f*(**Θ**) and the likelihood *L*(**Θ**,*η;****x***), the posterior distribution can be sampled by Monte Carlo Markov Chain simulations (MCMC). From these samples, we can obtain means and their 95% credible interval (CI) for the model parameters.

**Model fitting and estimation of model parameters**

Figure A1 show that within model variant 0, the three model parameters converge well. Model calibration also suggests that there should be about 2 (95% CI from 1 to 4) occult infections on Dec 1990. Figure A2 illustrates that the model fitting with observed epidemic data under model variant 0.

Within model variant I that is shown in Figure A3, the newly-included parameter: the proportion symptomatic (*p*S), appears not to well converge. The similarity between posterior and prior distributions implies that the observed epidemic data of symptom onset dates cannot provide any information about this parameter. However, *R*0 converges well and has nearly the same estimates as when *p*S is fixed at 85%.

Within model variant II that is shown in Figure A4, the two newly-included parameters: *p*S and *ρ* appear not to well converge. The similarity between their posterior and prior distributions implies that the observed epidemic data cannot provide any information about these two parameters. The posterior distribution of *ρ* affects that of growth rate and its associated parameters *β* and *R*0. Though not being converged as a bell-shaped curve, these parameters do converge to some restrictive range, respectively. Under model variant II, the estimate of *R*0 is 2.09[1.40,2.29] when *P*L:*P*OI:*P*SI =14:14:7. It becomes 2.16[1.48,3.00] if *P*L:*P*OI:*P*SI =18:15:7.

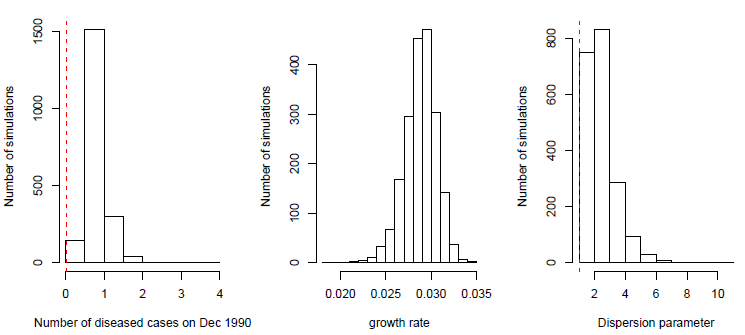
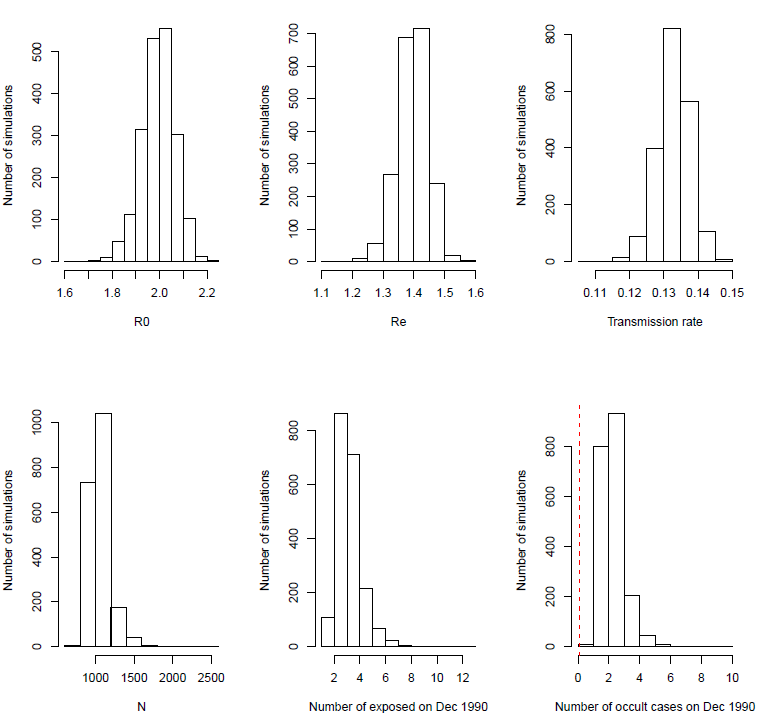


Figure A1 Posterior distribution of model parameters under model variant 0. The red vertical lines represent the lower and upper bounds of uniform priors (if the red vertical lines were not shown, it indicates that the bounds are beyond the range).

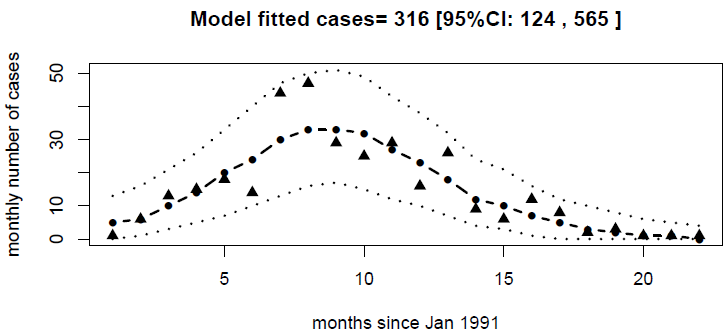


Figure A2 Model fitting to the observed symptom onset dates of 330 MSM patients under model variant 0. Triangles represent observed data, dark dashed line represent the median of model predictions, thin dashed lines the 95% CIs.

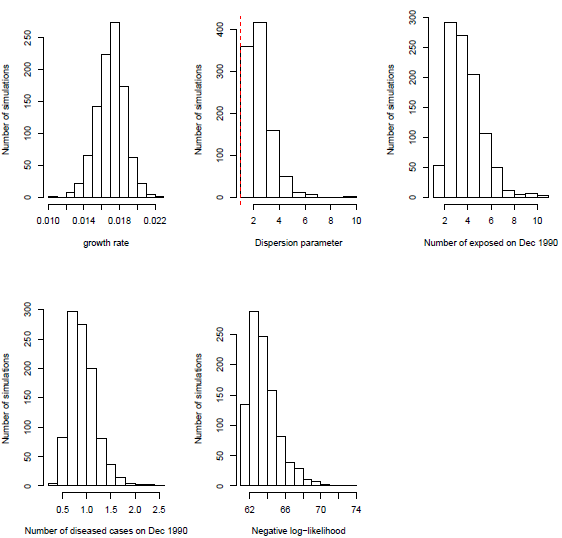
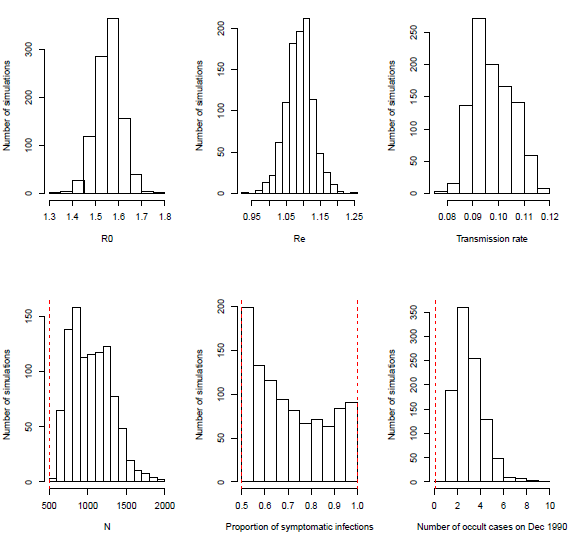


Figure A3 Posterior distributions of model parameters under model variant I. The proportion symptomatic (*p*S) appears not to converge. Model fitting to observed data is very similar to that shown in Figure A2. The red vertical lines represent the lower and upper bounds of uniform priors (if the red vertical lines were not shown, it indicates that the bounds are beyond the range).

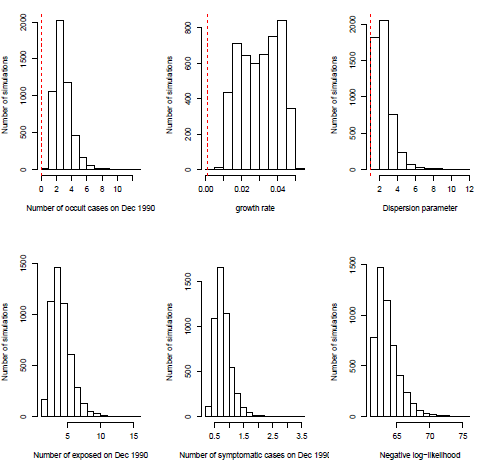
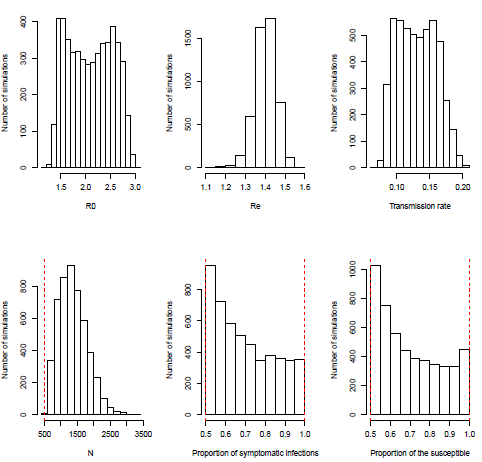


Figure A4 Posterior distributions of model parameters under model variant II. Both the proportion symptomatic (*p*S) and the initial proportion immune to HAV (*ρ*) appear to not converge. Model fitting to observed data is very similar to that shown in Figure A2. The red vertical lines represent the lower and upper bounds of uniform priors (if the red vertical lines were not shown, it indicates that the bounds are beyond the range) .

**Simulated outbreaks:**

The stochastic version of the model system (1) can be approximated by 13 discrete events listed in Tale A1. In Figure A5 the results of outbreak analyses for model variant 0 using MSM outbreak data are shown. The similar results for model variant I and II. As illustrated in Figure 3, these results show that our predicted outbreak probabilities are greatly decreased compared to that of Regan *et al*. (2016).

Table A1 List of events of stochastic version of equation (1) using Gillespie (1977) Methodology

|  |  |  |
| --- | --- | --- |
| Event | change | rate |
| Entry to susceptible MSMs | S→S+1 | *αN* |
| Exit from susceptible MSMs | S→S-1 | *αS* |
| Exit from exposed MSMs | E→E-1 | *αE* |
| Exit from occult infected MSMs | IO → IO-1 | *αIO* |
| Exit from asymptomatically infectious MSMs | IA → IA-1 | *αIA* |
| Exit from symptomatically infectious MSMs | IS → IS-1 | *αIS* |
| Exit from recovered and immune MSMs | R → R-1 | *αR* |
| New infection | E→E+1; S→S-1 | *βS*(*I*O+*I*A)/*N* |
| Removal from E to IO | IO → IO+1; E→E-1 | *σE* |
| Removal from IO to IA | IA → IA+1; IO → IO-1 | (1-*p*S) *γ*1*I*O |
| Removal from IO to IS | IS → IS+1; IO → IO-1 | *p*S*γ*1*I*O |
| Removal from IA to R | R → R+1; IA → IA-1 | *γ*2*I*A |
| Removal from IS to R | R → R+1; IS → IS-1 | *γ*2*I*S |

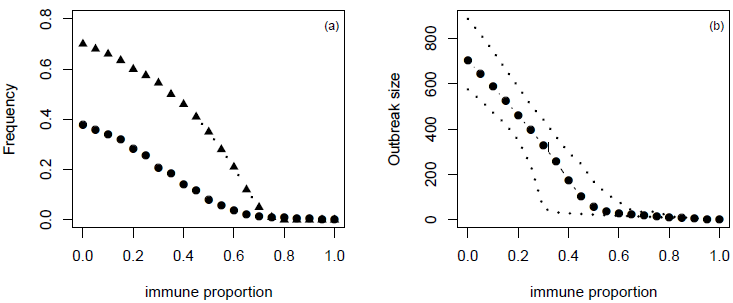


Figure A5 (a) outbreak probability as a function of the immune proportion, (b) the median size and 95%CI of such outbreaks under model variant 0 which was calibrated using MSM outbreak data. In panel (a) triangles represent the outbreak probability predicted by Regan et al (2016).

**Results of analysing the total outbreak**

In the above analysis, the model was fitted to cases reported in MSM (n=330). Among the further 240 cases (see Table 1 of Regan *et al* (2016)), only 53 cases were reported in women; Regan *et al* (2016) argued that a proportion of these 240 cases were also MSM. As in Regan *et al* (2016), the model variant III was also fitted to the total outbreak (n=570). The model fitting is illustrated in Figure A6. The posterior distributions of model parameters are quite similar to that discussed above for the outbreak of n=330 MSM cases. The summary of parameter estimates are listed in Table A2. Only population size and initial number of occult infection estimated to be different between the outbreak of 330 MSM cases and the outbreak of total 570 cases, which is understandable. Although the median of *R*0 for the outbreak of total 570 cases is slightly larger than that for the outbreak of 330 MSM cases, their 95% CI are nearly the same. The results of outbreak analyses for model variant III which was calibrated using the total outbreak data are shown in Figure A7. As illustrated in Figure 3 in the main text, Figure A7 show that under the different source data for model calibration, our predicted outbreak probabilities are greatly decreased compared to that of Regain (2016).

Table A2 Priors and posteriors distribution of model parameters under model variant III. For the convenience of comparison, the estimates (i.e. median and 95%CI) are shown for both outbreaks of 330 MSM cases and of total 570 cases. Here U stands for the uniform distribution.

|  |  |  |  |
| --- | --- | --- | --- |
| model parameters | priors | posterior | |
| Outbreak of 330 cases | Outbreak of 570 cases |
| *N* | U[500,10000] | 1274[676,2277] | 2346[1273,4083] |
| *P*L | U[10,18] | 13.9[10.1,17.9] | 13.9[10.1,17.8] |
| *P*OI | U[10,18] | 14.6[10.2,17.9] | 14.6[10.2,17.9] |
| PSI | U[3,11] | 6.9[3.2,10.9] | 7.0[3.1,10.8] |
| *p*S | U[0.5,1.0] | 0.68[0.51,0.98] | 0.66[0.51,0.98] |
| *ρ* | U[0.5,1.0] | 0.70[0.50,0.99] | 0.66[0.50,0.98] |
| *I*o(0) | U[0.1,20] | 2.72[1.20,6.07] | 4.16[1.90,8.62] |
| *Ψr* | U[0.01,0.10] | 0.029 [0.012,0.049] | 0.031[0.013,0.049] |
| *β* | – | 0.13[0.077,0.21] | 0.13[0.077,0.21] |
| *R*0 | – | 2.02[1.38,2.89] | 2.11[1.39,2.85] |
| *R*e | – | 1.41[1.27,1.57] | 1.39[1.27,1.54] |
| *η* | U[1.01,40] | 2.23[1.22,4.95] | 3.12[1.69,6.71] |

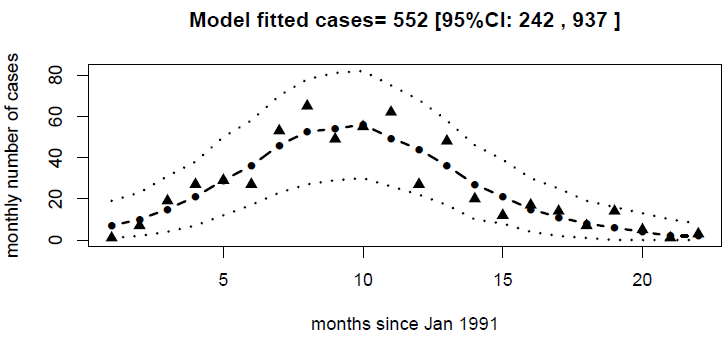


Figure A6 Model fitting to the observed symptom onset dates of 570 patients under model variant III. Triangles represent observed data, dark dashed line represent the median of model predictions, thin dashed lines the 95% CIs.

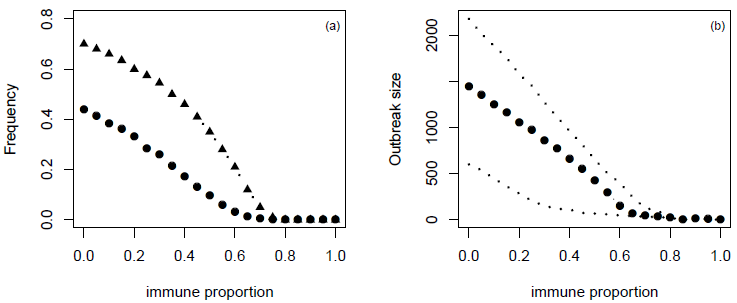
****

Figure A7 (a) outbreak probability as a function of the immune proportion, (b) the median size and 95%CI of such outbreaks under model variant III which was calibrated using the total outbreak data. In panel (a) triangles represent the outbreak probability predicted by Regan *et al* (2016).

**Effect of behaviour response during the outbreak**

When a large outbreak of hepatitis A occurs such as the one we examined in this study, local health agency must have implemented some interventions and health education which might decrease the contact rates between MSM people and hence reduce the transmissibility. To model this, we assume that transmission rate changes after a turning time point *t*c as

(7)

Here *w* is a factor introduced to describe the reduction in contact rates between MSM. This certainly simplifies the actual situations because such changes in contact rates are more likely to take place gradually rather suddenly from a time point. The transmission model equations are the same as equations (1) except replacing the constant *β* with a time-varying *β*(*t*) as in equation (7). To illustrate the effect of behaviour response during the outbreak, the model variant III is used with two additional parameters *t*c and *w*.

The results of analysis are listed in Table A3. It suggests that the contact and transmission rate is very like to reduce about 30% from day 213 (i.e., July 1991). The introduction of the varying *β*(*t*) only greatly alters the estimated value of size of MSM population (*N*) involved in the outbreak: without reduction it is 1270; with reduction it increases to 2811. However, this hardly alters the estimate of transmissibility and *R*0 still has a median 2.02 and a slightly changed 95CI: [1.36, 2.96]. The values of DIC listed in Table A3 for the two situations show that the inclusion of varying transmission rate cannot improve the model fitting.

Table A3 Priors and posteriors distribution of model parameters under model variant III without and with behaviour response. The median and 95%CI are shown for posteriors. The values of *t*c are the days from the onset of the outbreak.

|  |  |  |  |
| --- | --- | --- | --- |
| model parameters | priors | posterior | |
| Constant transmissibility | Varying transmissibility |
| *N* | U[500,10000] | 1274[676,2277] | 2811[882,9721] |
| *P*L | U[10,18] | 13.9[10.1,17.9] | 14.2[10.1,17.9] |
| *P*OI | U[10,18] | 14.6[10.2,17.9] | 14.8[10.2,17.9] |
| PSI | U[3,11] | 6.9[3.2,10.9] | 7.0[3.1,10.9] |
| *p*S | U[0.5,1.0] | 0.68[0.51,0.98] | 0.69[0.51,0.98] |
| *ρ* | U[0.5,1.0] | 0.70[0.50,0.99] | 0.7[0.51,0.99] |
| *I*o(0) | U[0.1,20] | 2.72[1.20,6.07] | 2.27[0.8,5.58] |
| *Ψr* | U[0.01,0.10] | 0.029 [0.012,0.049] | 0.029[0.12,0.05] |
| *t*c | U[100,330] | – | 213[112,301] |
| *w* | U[0.3,1.0] | – | 0.71[0.53,0.98] |
| *β* | – | 0.13[0.077,0.21] | 0.12[0.076,0.21] |
| *R*0 | – | 2.02[1.38,2.89] | 2.02[1.36,2.96] |
| *R*e | – | 1.41[1.27,1.57] | 1.41[1.23,1.70] |
| *η* | U[1.01,40] | 2.23[1.22,4.95] | 2.13[1.23,4.51] |
| DIC |  | 133.4 | 133.2 |