

A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data

S. Cauchemez^{1,2,*}, F. Carrat^{1,2,3}, C. Viboud^{1,4}, A. J. Valleron^{1,2,3}
and P. Y. Boëlle^{1,2,3}

¹INSERM U444, Paris, France

²Université Pierre & Marie Curie, Paris, France

³Assistance Publique—Hôpitaux de Paris, Paris, France

⁴Fogarty International Center, National Institute of Health, Bethesda, U.S.A.

SUMMARY

We propose a transmission model to estimate the main characteristics of influenza transmission in households. The model details the risks of infection in the household and in the community at the individual scale. Heterogeneity among subjects is investigated considering both individual susceptibility and infectiousness. The model was applied to a data set consisting of the follow-up of influenza symptoms in 334 households during 15 days after an index case visited a general practitioner with virologically confirmed influenza.

Estimating the parameters of the transmission model was challenging because a large part of the infectious process was not observed: only the dates when new cases were detected were observed. For each case, the data were augmented with the unobserved dates of the start and the end of the infectious period. The transmission model was included in a 3-levels hierarchical structure: (i) the observation level ensured that the augmented data were consistent with the observed data, (ii) the transmission level described the underlying epidemic process, (iii) the *prior* level specified the distribution of the parameters. From a Bayesian perspective, the joint *posterior* distribution of model parameters and augmented data was explored by Markov chain Monte Carlo (MCMC) sampling.

The mean duration of influenza infectious period was estimated at 3.8 days (95 per cent credible interval, 95 per cent CI [3.1,4.6]) with a standard deviation of 2.0 days (95 per cent CI [1.1,2.8]). The instantaneous risk of influenza transmission between an infective and a susceptible within a household was found to decrease with the size of the household, and established at 0.32 person day⁻¹ (95 per cent CI [0.26,0.39]); the instantaneous risk of infection from the community was 0.0056 day⁻¹ (95 per cent CI [0.0029,0.0087]). Focusing on the differences in transmission between children (less than 15 years old) and adults, we estimated that the former were more likely to transmit than adults (*posterior* probability larger than 99 per cent), but that the mean duration of the infectious period was similar in children (3.6 days, 95 per cent CI [2.3,5.2]) and adults (3.9 days, 95 per cent CI [3.2,4.9]). The *posterior* probability that children had a larger community risk was 76 per cent and the *posterior* probability that they were more susceptible than adults was 79 per cent. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: infectious diseases; influenza; household transmission; Bayesian inference; Markov chain Monte Carlo methods

*Correspondence to: Simon Cauchemez, INSERM U444, 27 Rue Chaligny, 75571, Paris Cedex 12, France.

†E-mail: simon.cauchemez@u444.jussieu.fr

1. INTRODUCTION

Household data have received particular attention in the study of influenza transmission, because they provide detailed information on the dynamics of infection within well-defined clusters of individuals. In such studies, a common design is to determine the serological statuses of all individuals in the household before and after an epidemic period, to allow the determination of who was infected. Using such data, Longini *et al.* [1, 2] proposed to estimate the probability of being infected, by an infective from the household or in the community, using a probabilistic model describing the total number of infections in a household at the end of an epidemic. Extensions to this approach made it possible to take into account heterogeneity in susceptibility or in infectiousness [3, 4].

In winter 1999–2000, we conducted a study comprising 334 households where the presence/absence of influenza symptoms was reported daily during 15 days after inclusion of an index case [5]. The order in which clinical cases occurred was therefore observed, and a substantial part of the available information would be disregarded in an analysis using the models described above. In particular, it is expected that these data would allow simultaneous estimation of the duration of the infectious period and the instantaneous risks of infection.

If the start and the end of the infectious period were available for each infective, the estimation of these parameters would be possible based on the likelihood of the data [6]. Unfortunately, here, only the dates of occurrence of symptoms were documented, and the epidemic was therefore only partially observed.

A practical approach consists of imputing dates for the start and the end of the infectious period for each case, compatible with the observation. Conditional on these augmented data, the likelihood of the epidemic data is readily available, but since there is no unique way to choose the augmented data given the observation, a systematic exploration of the augmented data is necessary for inference. Markov chain Monte Carlo (MCMC) sampling has proved useful in such problems [7–9].

For example, with data consisting of longitudinal indicators of carriage of *S. pneumoniae*, Auranen *et al.* [9] augmented the data with the unobserved dates of colonization/decolonization and were able to estimate the duration of carriage and the instantaneous risks of infection, using MCMC sampling.

An alternative approach to the analysis of such household data was proposed earlier by Rampey *et al.* [10], where, rather than augmenting the data, the authors fixed the distribution of the duration of the infectious period. This method allowed estimation of the probabilities of transmission between individuals in the same household and from the community, but was by construction unfit for estimating the duration of the infectious period.

In this paper, we aimed at estimating simultaneously the duration of the infectious period and the instantaneous risks of infection. We also investigated several hypotheses concerning the dynamics of transmission. More precisely, we examined the dependence of the risk of infection on the density of infectives, the distribution of the duration of the infectious period, and the role of children in the infectious process, e.g. whether children are more susceptible than adults; are more at risk of infection from the community than adults; are infectious for a longer duration than adults; are more infectious than adults. Therefore, we developed a 3-levels hierarchical model to analyse the data, and resorted to data augmentation with the unobserved dates of the start and the end of the infectious period for each case.

We present the data in Section 2, describe the model and MCMC sampling in Sections 3 and 4. Results and model validation are presented in Sections 5 and 6, and discussed in Section 7.

2. DATA

The Epigrippe study, described elsewhere in detail [5], took place during the 1999–2000 influenza season in France. In short, 946 households were recruited for follow-up by 161 general practitioners. Criteria for inclusion of a household were (i) visit of a member of the household to a general practitioner with a history of fever ($\geq 38^{\circ}\text{C}$) in the last 48 h and respiratory signs; (ii) at least another member in the household; (iii) the consulting patient was the first case in the household and was not hospitalized for his/her illness, and (iv) all patients agreed to participate in the study. The consulting patient was referred to as the index case of the household. In all index cases, nasal swabs were obtained at the inclusion visit and the presence of influenza virus was probed with an immunofluorescence test and/or viral culture and/or PCR. The circulating viruses during this epidemic were A/Sydney/5/97 (H3N2) and A/Moscow/10/99 (H3N2).

After inclusion, each household received a standardized report form where social and demographic characteristics were reported, as well as the daily time course of symptoms, medications, general practitioner visits for all household members during the 15 days following recruitment of the index case. For each member of the household and for each day, clinical influenza was defined as the presence of fever or feverishness, or at least 2 of the following signs: sore throat, headache, stiffness or myalgias, fatigue, cough, nasal congestion or rhinorrhea or sneezing.

Among the 946 index cases enrolled in the study, 510 were tested positive for the influenza virus. Follow-up information was obtained in 334 (65 per cent) of the concerned households. The data for analysis therefore consisted in 334 index cases (and households) and 790 contacts. We defined the contact cases as the contacts that had clinical influenza for at least 1 day. There were 350 such contact cases. Figure 1 presents the number of households for analysis depending on household size and on the number of children (subjects younger than 15 years old) in the household.

3. MODEL SPECIFICATION

A hierarchical description of the data was adopted, to take into account the natural household clustering, and the underlying epidemic process. The underlying epidemic process was cast into the classical SIR model, where a (S)usceptible individual may become (I)nfected and is ultimately (R)ecovered [11].

In a household f of size n_f , the presence/absence of clinical influenza for subject $i \in \{1, \dots, n_f\}$ at day $j \in \{0, \dots, 14\}$ was observed and denoted by the indicator variable Y_{ij}^f ($Y_{ij}^f = 1$ if clinical influenza was observed, else $Y_{ij}^f = 0$). The group of individuals reporting at least 1 day with clinical influenza during follow-up was denoted I^f , and the remaining members S^f . The first day with clinical influenza for subject $i \in I^f$ was denoted Z_i^f . We considered that each individual i in I^f had been infected with the influenza virus, and

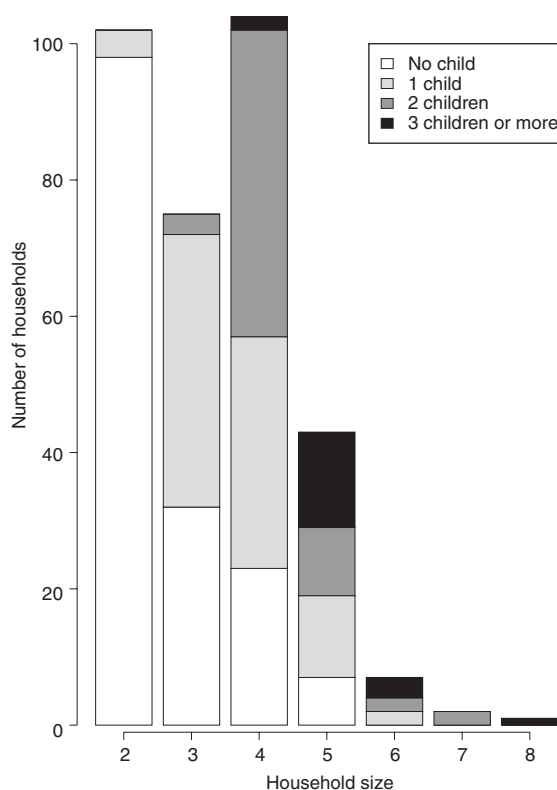


Figure 1. Number of households for analysis depending on household size and on the number of children (subjects younger than 15 years old) in the household.

defined two unobserved variables, v_i^f and ψ_i^f ($v_i^f < \psi_i^f$) corresponding to the start and the end of the infectious period for individual i . These dates were considered in continuous time, to always provide an unambiguous ordering of occurrence of infections in a household. Last, we defined $I^f(t) = \{i \in I^f, v_i^f < t \leq \psi_i^f\}$ the group of infectives just before time t . $Y^f = (Y_{ij}^f, i \in \{1, \dots, n^f\}, j \in \{0, \dots, 14\})$ corresponded to all observations in household f , and Y in all households. In the same way, we defined v^f , ψ^f , v and ψ .

Denoting model parameters as θ , the joint density of observations, unobserved variables, and parameters was

$$P(Y, v, \psi, \theta) = P(Y|v, \psi)P(v, \psi|\theta)P(\theta) \quad (1)$$

where $P(Y|v, \psi) = \prod_f P(Y^f|v^f, \psi^f)$ and $P(v, \psi|\theta) = \prod_f P(v^f, \psi^f|\theta)$.

Following Auranen *et al.* [9], the factors on the right-hand side of equation (1) are referred to as the observation level, the transmission level and the *prior* level, respectively. This formulation implied that (i) observations and underlying infectious processes were independent across households, (ii) conditional on the realization of the underlying infectious process, observations in a household were independent of the model parameters.

The structure of each hierarchical level is detailed below. The observation and transmission levels are presented for one household f , where we omit the indicator f for the sake of clarity.

3.1. Observation level

The role of this level was to ensure agreement between the augmented data (v, ψ) and the observed data Y . The incubation period, defined as the duration between infection and the start of symptoms [12], ranges between 1 and 3 days for influenza [13–17]. There is no definite evidence on the relationship between the end of the infectious period and that of the symptoms. Therefore, we considered that the augmented data were compatible with the observed data if the start of infectious period v fell between 1 and 3 days before the start of clinical influenza Z , and if the end of infectious period ψ happened after v . As v is continuous and Z discrete, we wrote

$$P(Y|v, \psi) = \prod_{i \in I} \mathbf{1}\{Z_i - 3 < v_i < Z_i \text{ and } v_i < \psi_i\}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function.

3.2. Transmission level

This level described influenza transmission within a household, assuming the dates v and ψ known. In a household of size n , for a subject s susceptible just before t , the instantaneous risk of infection at time t was

$$\lambda_s(t) = \alpha_s + \varepsilon_s \sum_{i \in I(t)} \frac{\beta_i}{n}$$

where α_s was the instantaneous risk of infection from the community, which may depend on the characteristics of susceptible s ; $\varepsilon_s \beta_i n^{-1}$ was the instantaneous risk of infection due to infective i for susceptible s . More precisely, ε_s measured the susceptibility to infection of individual s , and β_i measured the ability of infective i to infect others. To investigate the dependence of the risk on the density of infectives, we also considered a model in which the contribution of infective i in a household of size n was $\beta_i n^{-\eta}$, instead of $\beta_i n^{-1}$, and estimated η along with the other parameters.

The duration of infectious period $\psi_i - v_i$ for infective i was taken as a gamma distribution with mean μ_i , standard deviation σ_i and density d_{μ_i, σ_i} .

Therefore, conditional on the date of the first infection v_1 , the likelihood of the process in the household was

$$P(v, \psi | \theta) = \prod_{i \in I} d_{\mu_i, \sigma_i}(\psi_i - v_i) \prod_{i \in I - \{1\}} \lambda_i(v_i) e^{-\int_{v_1}^{v_i} \lambda_i(t) dt} \prod_{s \in S} e^{-\int_{v_1}^{\psi_s} \lambda_s(t) dt}$$

where $I - \{1\}$ was the group of infectives I without the first infected. Note that subject 1 did not necessarily correspond to the index case of the household, but to the subject with minimum v .

Two summary quantities adopted in previous works on infectious transmission in households [2, 4] were calculated from the parameters in the model. The community probability of

infection (CPI), defined as the probability for subject i to be infected from the community during the 15-days follow-up, was

$$\text{CPI}_i = 1 - e^{-15\alpha_i}$$

The secondary attack rate (SAR), defined as the probability that infective i infects susceptible s in a household of size n , was

$$\text{SAR}_{i \rightarrow s}(n) = 1 - \int_0^{+\infty} \exp\left(-\varepsilon_s \frac{\beta_i}{n} t\right) d_{\mu_i, \sigma_i}(t) dt$$

3.3. Investigating the role of children

To investigate the specific role of children in the infectious process, we allowed the parameters to differ for children (subjects younger than 15 years old) and adults. The corresponding parameters were indexed by C for children and A for adults.

We therefore investigated the parameters (α_C, α_A) for community infection, (β_C, β_A) for household infection, (μ_C, μ_A) for the mean duration of the infectious period. The standard deviation of the infectious period was kept identical in the two groups ($\sigma_C = \sigma_A = \sigma$). For the susceptibility to infection according to age, we defined adults as the reference category. Therefore, we wrote $\varepsilon_A = 1$ and estimated ε_C , the susceptibility ratio between children and adults exposed to the same source of infection in the household.

Using this parameterization, five hypotheses about the role of children were formulated for investigation:

- Hypothesis $H1$: children are more susceptible than adults, i.e. $\varepsilon_C > 1$.
- Hypothesis $H2$: community risk of infection is larger for children than for adults, i.e. $\alpha_C > \alpha_A$.
- Hypothesis $H3$: household risk of infection is larger when the infective is a child than when he/she is an adult, i.e. $\beta_C > \beta_A$.
- Hypothesis $H4$: children are infectious for a longer duration than adults, i.e. $\mu_C > \mu_A$.
- Hypothesis $H5$: SAR is larger when the infective is a child than when he/she is an adult. SAR depends on both the household size and the ages of the infective and the susceptible. For household size n in $\{2, \dots, 5\}$, we compared the probabilities of infection associated with infectious child/adult for a susceptible adult, i.e. $H5(n)$: $\text{SAR}_{C \rightarrow A}(n) > \text{SAR}_{A \rightarrow A}(n)$. Subsequently, we will write $\text{SAR}_C(n) = \text{SAR}_{C \rightarrow A}(n)$ and $\text{SAR}_A(n) = \text{SAR}_{A \rightarrow A}(n)$. SAR combines information on the instantaneous risk of infection β and the mean duration of infectious period μ , and therefore allowed investigation of whether children were 'globally' more infectious than adults.

3.4. Prior level

Independent *prior* distributions were chosen for μ , σ , α , β , ε and η . Since the mean duration of influenza infectious period is reportedly in the range 2–5 days [18–20], we adopted a gamma distribution with mean 3 and standard deviation 2 as *prior* distribution for both μ and σ . There is much less information available for α and β . We therefore adopted the exponential distribution $\text{Exp}(0.001)$ for both parameters to reflect the lack of information.

For the susceptibility ratio ε_C , we specified the prior such that $\log(\varepsilon_C)$ is logistic with scale parameter s . This prior satisfies the invariance condition that the ratio (adult susceptibility/child susceptibility) has the same prior as the ratio (child susceptibility/adult susceptibility). In particular, it gives equal probabilities to ε_C being larger or smaller than 1. We took $s=1$ so that there would be a large prior 95 per cent credible interval $([0.025,40])$. Last, the *prior* distribution for η was uniform on the range $[-3,3]$. When the parameters μ , α , β were allowed to differ between children and adults, the same *prior* distributions were selected for both categories.

4. MCMC SAMPLING

We used a Markov chain Monte Carlo for estimation [21]. In this approach, a Markov chain is constructed so that its stationary distribution is the *posterior* distribution $P(\theta, v, \psi|Y)$ of the parameters and the augmented data given the observed data.

The chain was started with augmented data that respected observation level and overdispersed parameters. For each infective i in household f , the start of the infectious period v_i^f was drawn in uniform distribution $U[Z_i^f - 3, Z_i^f]$ and the duration of the infectious period $\psi_i^f - v_i^f$ was drawn in $U[0,20]$. α and β were drawn in $U[0,1]$, μ and σ in $U[0,10]$, ε_C and η in their *priors*.

We performed a single-component Metropolis–Hastings sampling [21]. At each iteration, the following moves were applied sequentially: (a) resampling the model parameters α , β and ε ; (b) resampling the model parameters μ and σ ; (c) resampling one duration of infectious period per household; (d) resampling one start of infectious period per household.

For moves a–c, we performed random-walk Metropolis sampling [21]. If the current value of the component was ρ , a new value ρ^* was generated so that $\log(\rho^*) = \log(\rho) + \delta u$ with u drawn from the normal distribution $N(0,1)$. We specified $\delta=0.1$ for moves a, $\delta=0.03$ for moves b and $\delta=1$ for moves c for better mixing. For moves d, we performed independence Metropolis sampling [21]. For each household f , an infective i was randomly selected and a new candidate v_i^{f*} was drawn from $U[Z_i^f - 3, Z_i^f]$. The end of the infectious period was modified so that the duration of the infectious period did not change, i.e. $\psi_i^{f*} - v_i^{f*} = \psi_i^f - v_i^f$.

We performed 200 000 iterations for each run of the MCMC algorithm. The first 5000 were discarded as the burn-in period. The output was then recorded one every 10 iterations to constitute a sample from the *posterior* distribution. With a Pentium III processor, one procedure of estimation took 30 min. The convergence of the MCMC was tested with the Gelman–Rubin criterion (GRC) [22]: 5 chains were run with different starts; the GRC was estimated for each parameter and for the log-likelihood. A value close to 1 (under 1.1) was a sign for convergence.

From the MCMC samples, we estimated the *posterior* means, *posterior* standard deviation and 95 per cent equal-tailed credible intervals of parameters, and the *posterior* probabilities of hypothesis H_1 – H_5 . For $i=1, \dots, 5$, the Bayes factor of hypothesis H_i against \bar{H}_i was calculated

$$\text{BF}_i = \frac{P(H_i|Y) P(\bar{H}_i)}{P(\bar{H}_i|Y) P(H_i)}$$

where $P(H_i)$ and $P(H_i/Y)$ are, respectively, the *prior* and *posterior* probabilities for H_i . The Bayes factors results are therefore inherently dependent on the choice of priors for the model parameters. Here, for parameters α , β and μ , the same *priors* were chosen for adults and children, so that the *prior* probability that children parameter is larger than adults parameter was 0.5 for $H2-H5$. Likewise, the *prior* distribution for ε_C was specified so that $P(\varepsilon_C > 1) = P(\varepsilon_C < 1)$. Therefore, for $i = 1, \dots, 5$, the Bayes factor reduced to the ratio of the *posterior* probabilities of H_i and \bar{H}_i

$$\text{BF}_i = \frac{P(H_i/Y)}{1 - P(H_i/Y)}$$

Following Jeffrey [23], a level of evidence was associated to a Bayes factor: none ($\log_{10} \text{BF} < 0.0$); poor ($0.0 < \log_{10} \text{BF} < 0.5$); substantial ($0.5 \leq \log_{10} \text{BF} < 1.0$); strong ($1.0 \leq \log_{10} \text{BF} < 2.0$); decisive ($2.0 \leq \log_{10} \text{BF}$). Simulations were performed to determine the typical values of the Bayes factors under various scenarios.

We used the MCMC samples drawn from the joint *posterior* distribution of (μ, σ) , to draw a sample from the *posterior* distribution of the duration of the infectious period. This step allowed to estimate *posterior* quantiles of the duration of the infectious period.

5. RESULTS

5.1. Parameters for influenza transmission

We first estimated the model assuming that there was no difference between children and adults concerning transmission. The susceptibility parameter ε was fixed at 1 for all subjects; therefore only 4 parameters were estimated, namely $(\alpha, \beta, \mu, \sigma)$. Figure 2 presents the MCMC output, the *prior* and *posterior* densities of the parameters. Visual inspection indicates convergence of the MCMC algorithm, in agreement with the GRC (< 1.01 for all parameters). For the four parameters, the *posterior* distributions covered narrow intervals, that were shorter than those defined by the *prior* distributions. In particular, we obtained an informative *posterior* description for α and β from the relatively uninformative $\text{Exp}(0.001)$ *prior*.

Table I presents the *posterior* means (*posterior* standard deviation) and 95 per cent equal-tailed credible intervals for model parameters, SAR and CPI. Estimates for duration parameters were $\mu = 3.8[3.1, 4.6]$ days and $\sigma = 2.0[1.1, 2.8]$ days, yielding a 95 per cent credible interval for the duration of the infectious period of $[0.8, 8.6]$. Depending on the size of the household n , the instantaneous risk of infection in the community was 11 ($n = 5$) to 29 ($n = 2$) times smaller than the instantaneous risk of infection from an infective in the same household. This was also reflected in the value of the CPI, which was always smaller than the SAR. The SAR decreased with the size of the household, from 0.43 $[0.39, 0.48]$ in households of size 2 to 0.21 $[0.18, 0.24]$ in households of size 5, showing that transmission from one infective to a susceptible was more effective in small households.

There was little *posterior* correlation between the risks of infection α and β , $\text{cor}(\alpha, \beta) = 0.01$. But, as expected, these parameters were inversely related to the mean duration of the infectious period μ , $\text{cor}(\alpha, \mu) = -0.44$ and $\text{cor}(\beta, \mu) = -0.64$. There was little correlation between the duration parameters, $\text{cor}(\mu, \sigma) = 0.19$.

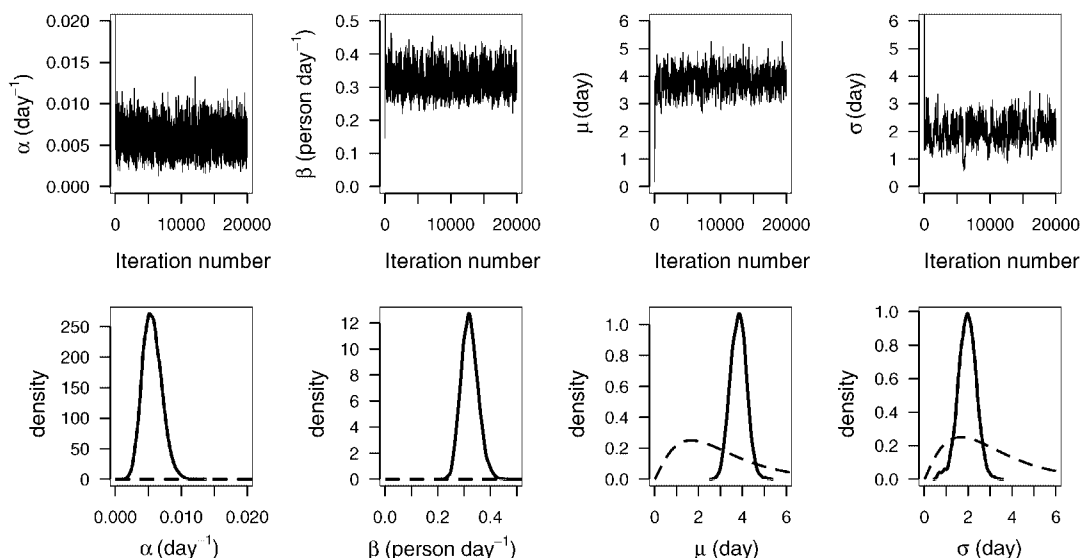


Figure 2. MCMC output, *prior* and *posterior* distributions for the community risk of infection (α), the household risk of infection (β), the mean duration of infectious period (μ) and the standard deviation of infectious period (σ). Dotted lines represent *priors*.

Table I. *Posterior* means (*posterior* standard deviations) and 95 per cent credible intervals (95 per cent CI) for model parameters, secondary attack rates (SAR) and community probability of infection (CPI). SAR is a function of the size of the household.

	Overall		Children		Adults	
	Mean(SD*)	95 per cent CI	Mean(SD*)	95 per cent CI	Mean(SD*)	95 per cent CI
<i>Model parameters</i>						
Individual susceptibility, ε	—	—	1.15(0.19)	[0.81, 1.56]	1(—) [†]	—
Community risk, α (10^{-3} day $^{-1}$)	5.6(3.1)	[2.9, 8.7]	7.9(3.1)	[2.7, 14.8]	5.4(1.6)	[2.6, 8.8]
Household risk, β (person day $^{-1}$)	0.32(0.03)	[0.26, 0.39]	0.48(0.12)	[0.29, 0.75]	0.26(0.04)	[0.19, 0.33]
Infectious period, mean μ (day)	3.8(0.4)	[3.1, 4.6]	3.6(0.7)	[2.3, 5.2]	3.9(0.4)	[3.2, 4.9]
Infectious period, SD* σ (day)	2.0(0.4)	[1.1, 2.8]	1.9(0.4) [‡]	[1.0, 2.7] [‡]	1.9(0.4) [‡]	[1.0, 2.7] [‡]
<i>SAR and CPI</i>						
SAR(2)	0.43(0.02)	[0.39, 0.48]	0.52(0.05)	[0.43, 0.62]	0.38(0.03)	[0.32, 0.43]
SAR(3)	0.31(0.02)	[0.28, 0.36]	0.40(0.04)	[0.32, 0.49]	0.27(0.02)	[0.23, 0.32]
SAR(4)	0.25(0.02)	[0.22, 0.27]	0.33(0.04)	[0.25, 0.40]	0.22(0.02)	[0.18, 0.25]
SAR(5)	0.21(0.01)	[0.18, 0.24]	0.27(0.03)	[0.21, 0.34]	0.18(0.02)	[0.15, 0.21]
CPI	0.08(0.02)	[0.04, 0.12]	0.11(0.04)	[0.04, 0.20]	0.08(0.02)	[0.04, 0.12]

*Standard deviation.

[†]We defined adults as the reference group and specified $\varepsilon_A = 1$.

[‡]We assumed $\sigma_C = \sigma_A$.

Table II. *Prior* and *posterior* probabilities, Bayes factors for hypothesis $H1-H5$.

Hypothesis	Prior	Post	BF	\log_{10} BF	Evidence
$H1$: Children are more susceptible than adults $\varepsilon_C > 1$	0.50	0.79	3.7	0.57	Substantial
$H2$: Community risk is larger for children than for adults $\alpha_C > \alpha_A$	0.50	0.76	3.1	0.50	Substantial
$H3$: Children are infectious for a longer duration than adults $\mu_C > \mu_A$	0.50	0.32	0.48	-0.33	Poor (against)
$H4$: Household risk is larger with infectious child than with infectious adult $\beta_C > \beta_A$	0.50	0.98	43.4	1.63	Strong
$H5$: SAR is larger with infectious child than with infectious adult for $n \in \{2, \dots, 5\}$ $SAR_C(n) > SAR_A(n)$	0.50	>0.99	>249	>2.4	Decisive

5.2. The role of children in the transmission process

When the parameters were allowed to differ between children and adults, visual inspection of the MCMC output as well as the GRC (< 1.03 for each parameter) indicated convergence of the MCMC algorithm. Table I presents the *posterior* means (*posterior* standard deviation) and 95 per cent equal-tailed credible intervals for model parameters, SAR and CPI. Compared to the overall estimates, the only noticeable changes were in the parameters corresponding to instantaneous risks of infection: β_C was found twice larger in children than in adults, and community risk about 1 third larger in children. The *posterior* correlation between adults parameters was of the same order as the one observed between overall parameters. Children risks were positively correlated, $\text{cor}(\alpha_C, \beta_C) = 0.34$, and negatively correlated with children susceptibility, $\text{cor}(\alpha_C, \varepsilon_C) = -0.35$ and $\text{cor}(\beta_C, \varepsilon_C) = -0.47$. There was little correlation between children and adults parameters.

Table II presents *prior* and *posterior* probabilities, and Bayes factors for hypotheses $H1-H5$. There was substantial evidence that children were more susceptible than adults ($H1$: BF = 3.7) and that they had a larger community risk ($H2$: BF = 3.1). There was poor evidence that children were infectious for a shorter duration than adults ($H3$: BF = 0.48), but there was strong evidence that household risk was larger for an infectious child than for an infectious adult ($H4$: BF = 43.4). Irrespective of the size of the household, there was decisive evidence that secondary attack rates were larger when the infectious was a child ($H5$: BF > 249.0).

6. SIMULATION

6.1. Model checking

In order to check the model adequation to the data, we simulated epidemics in a community of households with the same size and structure as the original data, using an SIR epidemic model with parameters drawn from the *posterior* distributions (see appendix). Table III presents observed and expected distributions of the number of cases per household obtained from

Table III. Observed and expected distributions (mean [2.5 and 97.5 per cent quantiles]) of the number of cases per household. Expected distribution was derived from 5000 epidemic simulations with parameters drawn from the *posterior* distribution.

Household size	Distribution	Number of cases				
		1	2	3	4	5
2	Observed	53	49			
	Expected	53.5[43,64]	48.5[38,59]			
3	Observed	28	24	23		
	Expected	30.7[22,40]	22.6[15,31]	21.7[14,30]		
4	Observed	31	22	33	18	
	Expected	36.1[26,47]	25.0[16,34]	22.9[15,31]	20.0[12,29]	
5	Observed	11	11	10	7	4
	Expected	13.2[6,19]	9.0[4,15]	7.9[3,13]	7.3[3,13]	5.6[2,11]

5000 simulations. There was excellent agreement between the two distributions ($\chi^2 = 6.9$, 10 df, $P = 0.73$).

To check the dependence of the risk of infection on the density of infectives, we explored a model in which the contribution of infective i in a household of size n was $\beta_i n^{-\eta}$. *Posterior* mean and 95 per cent credible interval were $\eta = 0.84[0.46, 1.21]$, which confirmed that the formulation we adopted with $\eta = 1$ was close to optimal. We also estimated the parameters of the model where the risk of infection accrued by the same amount with each infective, irrespective of the size of the household, i.e. $\eta = 0$. *Posterior* distributions did not change much except of course for β ($0.085[0.069, 0.103]$), but the fit of the model worsened ($\chi^2 = 27.6$, 10 df, $P = 0.002$). More precisely, the number of cases was overestimated in large households and underestimated in small ones in the simulations. This was expected as the observed proportion of contact cases decreased with the size of the household (48, 47, 45 and 39 per cent in households of size 2, 3, 4, 5, respectively) when the model with $\eta = 0$ forced the risk of infection to increase with the size of the household.

Although the clinical assumption of short incubation time of influenza is well grounded, longer incubation times have been reported, even if such incubation times are more unusual. We therefore estimated the model with a less restrictive observation model: the incubation time could be in interval 1–10 days, but with the *prior* probability that it was in 1–3 days twice as large as the *prior* probability that it was in 4–10 days. No difference was found in the estimates. When the incubation time was assumed to be in 1–6 days: the estimate of the mean and standard deviation of the duration of the infectious period were roughly 20 h shorter ($\mu = 3.0[2.2, 3.9]$ days and $\sigma = 1.3[0.4, 2.2]$ days), while the estimate of the household risk increased by 25 per cent ($\beta = 0.40[0.31, 0.52]$ person day⁻¹); the community risk was unchanged ($\alpha = 0.0055[0.0031, 0.0083]$ day⁻¹). For larger intervals for the incubation time, we did not observe convergence of the MCMC algorithm.

We also investigated the distribution of the infectious period, by adopting a gamma distribution for the duration of the infectious period rather than the typical exponential distribution associated with SIR models. We found that the *posterior* probability that $\mu > \sigma$ was larger than 0.99, yielding decisive evidence that the distribution of the duration of the infectious period

Table IV. Mean (standard deviation) and biases of parameters *posterior* means estimated from 100 simulated data.

Parameter	Simulation value	15-days follow-up		1-month follow-up	
		Mean(SD*)	Bias(per cent)	Mean(SD*)	Bias(per cent)
Community risk, α (10^{-3} day $^{-1}$)	5.6	4.5(1.2)	-0.19	5.3(0.6)	-0.05
Household risk, β (person day $^{-1}$)	0.32	0.32(0.03)	0.0006	0.32(0.03)	0.01
Infectious period, mean μ (day)	3.8	3.9(0.3)	0.03	3.9(0.3)	0.02
Infectious period, SD* σ (day)	2.0	2.0(0.4)	-0.004	2.0 0.40	-0.006

*Standard deviation.

The empirical biases are also given.

was not exponential. We estimated the model when forcing the duration of the infectious period to have exponential distribution, i.e. fixing $\mu = \sigma$, and obtained $\mu = \sigma = 3.3[2.5, 4.1]$. The 95 per cent credible interval for the duration of infectious period was much larger (its size increased by 50 per cent): it was [0.9, 8.6] with a gamma distribution, and [0.1, 12.0] with an exponential distribution.

6.2. Performance in estimating influenza parameters

To investigate the bias in the estimation procedure, 100 epidemics were simulated in a community of households with the same size and structure as the original data (see appendix), using an SIR epidemic model with parameters set at known values. We varied the length of follow-up in the households from 15 days to 1 month. The *posterior* mean of each parameter was estimated for each simulated epidemic, and we report the mean and standard deviation of these values in Table IV for the two durations of households follow-up. For 15 days follow-up, the bias was smaller than 5 per cent for intra-household parameters (household risk β , mean μ and standard deviation σ of the duration of the infectious period), while the community risk α was 19 per cent under-estimated. When households follow-up was extended to 1 month, the empirical bias for α was reduced to 5 per cent.

6.3. Performance in determining the truth of hypotheses $H1-H4$

Five scenarios were considered to investigate the performance of our method to determine the truth of hypotheses $H1-H4$. In the first scenario $S0$, children and adults have the same parameters. The 4 other scenarios enabled us to investigate the performance of our method to detect changes in 1 parameter: $\varepsilon_C/\varepsilon_A = 1.5$ ($S1$), $\alpha_C/\alpha_A = 1.5$ ($S2$), $\mu_C/\mu_A = 1.5$ ($S3$) and $\beta_C/\beta_A = 1.5$ ($S4$). For each scenario, 100 epidemics were simulated in a community of households with the same size and age structure as the original data (see appendix). Table V presents the estimates based on simulated data.

Under $S0$, the median ratios $\varepsilon_C/\varepsilon_A$, μ_C/μ_A and β_C/β_A were in [0.95, 1.05], but the community risk median ratio α_C/α_A was 1.20. The \log_{10} BF distributions were approximately centered on 0. For each hypothesis investigated, there was at least 45 per cent probability to obtain poor evidence and 80 per cent probability to obtain poor to substantial evidence (in favour or against).

Table V. Simulation study of the relative susceptibility (ϵ_C/ϵ_A), community risk (α_C/α_A), duration of the infectious period (μ_C/μ_A), household risk (β_C/β_A) of children and adults; and of Bayes factors for hypotheses $H1-H4$. For five scenarios of transmission S0-S4, 100 epidemics were simulated in a community of households with the same size and age structure as the original data. For each scenario, the table gives: (1) the true value of the ratio between children and adults parameters, the median and [2.5 and 97.5 per cent] quantiles of the ratio of posterior means of children and adults parameters; (2) the distributions of the Bayes factors for hypotheses $H1-H4$.

Scenario	Ratio	True value	Estimate	Hypothesis investigated	Distribution of the Bayes factor $\log_{10} BF$							
					$] -\infty, -2]$	$] -2, -1]$	$] -1, -0.5]$	$] -0.5, 0.5]$	$] 0.5, 1]$	$] 1, 2]$	$] 2, +\infty[$	
					Decisive (against)	Strong (against)	Substantial (against)	Poor	Substantial	Strong	Decisive	
S0	ϵ_C/ϵ_A	1	0.98[0.73, 1.38]	$H1 : \epsilon_C > \epsilon_A$	0.00	0.14	0.19	0.47	0.14	0.05	0.01	
	α_C/α_A	1	1.20[0.34, 3.66]	$H2 : \alpha_C > \alpha_A$	0.00	0.05	0.12	0.48	0.19	0.15	0.01	
	μ_C/μ_A	1	0.96[0.70, 1.37]	$H3 : \mu_C > \mu_A$	0.00	0.10	0.29	0.45	0.11	0.05	0.00	
	β_C/β_A	1	1.05[0.69, 1.80]	$H4 : \beta_C > \beta_A$	0.00	0.01	0.13	0.61	0.12	0.12	0.01	
S1	ϵ_C/ϵ_A	1.5	1.45[1.09, 2.01]	$H1 : \epsilon_C > \epsilon_A$	0.00	0.00	0.00	0.05	0.05	0.37	0.53	
	α_C/α_A	1	1.10[0.37, 2.84]	$H2 : \alpha_C > \alpha_A$	0.00	0.03	0.12	0.61	0.17	0.05	0.02	
	μ_C/μ_A	1	0.99[0.75, 1.30]	$H3 : \mu_C > \mu_A$	0.01	0.03	0.23	0.64	0.06	0.02	0.01	
	β_C/β_A	1	1.04[0.65, 1.59]	$H4 : \beta_C > \beta_A$	0.01	0.04	0.12	0.58	0.17	0.08	0.00	
S2	ϵ_C/ϵ_A	1	1.04[0.71, 1.41]	$H1 : \epsilon_C > \epsilon_A$	0.01	0.12	0.11	0.50	0.15	0.10	0.01	
	α_C/α_A	1.5	1.54[0.57, 3.10]	$H2 : \alpha_C > \alpha_A$	0.00	0.02	0.03	0.40	0.26	0.20	0.09	
	μ_C/μ_A	1	0.99[0.69, 1.37]	$H3 : \mu_C > \mu_A$	0.00	0.08	0.16	0.60	0.11	0.04	0.01	
	β_C/β_A	1	1.05[0.61, 1.65]	$H4 : \beta_C > \beta_A$	0.00	0.07	0.17	0.52	0.16	0.07	0.01	
S3	ϵ_C/ϵ_A	1	0.96[0.69, 1.24]	$H1 : \epsilon_C > \epsilon_A$	0.02	0.15	0.16	0.52	0.12	0.02	0.01	
	α_C/α_A	1	1.15[0.44, 3.50]	$H2 : \alpha_C > \alpha_A$	0.01	0.01	0.17	0.51	0.17	0.10	0.03	
	μ_C/μ_A	1.5	1.33[0.95, 1.79]	$H3 : \mu_C > \mu_A$	0.00	0.01	0.00	0.19	0.19	0.35	0.26	
	β_C/β_A	1	1.15[0.74, 1.79]	$H4 : \beta_C > \beta_A$	0.00	0.03	0.07	0.46	0.21	0.18	0.05	
S4	ϵ_C/ϵ_A	1	1.00[0.74, 1.35]	$H1 : \epsilon_C > \epsilon_A$	0.01	0.09	0.14	0.50	0.16	0.10	0.00	
	α_C/α_A	1	1.01[0.38, 2.97]	$H2 : \alpha_C > \alpha_A$	0.00	0.05	0.19	0.55	0.12	0.08	0.01	
	μ_C/μ_A	1	0.99[0.68, 1.33]	$H3 : \mu_C > \mu_A$	0.03	0.10	0.15	0.54	0.12	0.05	0.01	
	β_C/β_A	1.5	1.55[1.04, 2.57]	$H4 : \beta_C > \beta_A$	0.00	0.00	0.01	0.10	0.18	0.40	0.31	

Table VI. Effect of the *prior* for the susceptibility ratio ε_C on the *posterior* distribution of ε_C and on the Bayes factors. The *prior* for ε_C is such that $\log(\varepsilon_C)$ is logistic with scale parameter s .

	$s = 0.1$	$s = 0.2$	$s = 0.3$	$s = 1$	
Mean[95 per cent CI] for ε_C		Informative* <i>priors</i> for (μ, σ)			
<i>prior</i>	1.0[0.69,1.44]	1.0[0.48,2.1]	1.0[0.33,3.00]	1.0[0.025,40]	
<i>posterior</i>	1.07[0.85,1.35]	1.12[0.82,1.47]	1.14[0.82,1.51]	1.15[0.81,1.56]	
\log_{10} BF		Informative*/flat† <i>priors</i> for (μ, σ)			
<i>H1: $\varepsilon_C > 1$</i>	0.39/0.41	0.49/0.51	0.53/0.55	0.57/0.60	
<i>H2: $\alpha_C > \alpha_A$</i>	0.65/0.62	0.57/0.54	0.54/0.51	0.50/0.47	
<i>H3: $\mu_C > \mu_A$</i>	-0.43/-0.30	-0.40/-0.27	-0.34/-0.23	-0.33/-0.22	
<i>H4: $\beta_C > \beta_A$</i>	1.97/1.71	1.85/1.64	1.73/1.54	1.63/1.54	

*Gamma distribution with mean 3 days and standard deviation 2 days.

†Exponential distribution $\text{Exp}(0.001)$.

Compared to the distributions under S0, the \log_{10} BF distributions under S1–S4 clearly indicated the parameter that was changed in the corresponding scenario. Under S1, the probability to obtain strong or decisive evidence in favour of hypothesis *H1* was 90 per cent, while it was only 6 per cent under S0. The same changes were observed for *H3* and *H4*: 61 per cent of strong or decisive evidence in favour of *H3* under S3 as opposed to 5 per cent under S0; 71 per cent of strong or decisive evidence in favour of *H4* under S4 as opposed to 13 per cent under S0. The change was smaller for *H2*: 26 per cent of strong or decisive evidence in favour of *H2* under S2 as opposed to 16 per cent under S0. Finally, in all scenarios, the estimated median ratios were close to the true value.

The Bayes factor estimated from the original data for *H4* was in the upper tail of its distribution under S0 (probability that a BF estimated under S0 is larger than the BF estimated from the original data: 0.04). In contrast, the Bayes factors estimated from the original data for *H1*, *H2* and *H3* were more compatible with S0 (probability that a BF estimated under S0 is larger than the BF estimated from the original data: 0.17, 0.35 and 0.56, respectively).

We also investigated the effect of changing the *prior* distributions on the Bayes factors. A particular concern was the long right tail of the *prior* for ε_C that might have some effect on the corresponding *posterior* probabilities. The Bayes factors were therefore calculated using narrower *prior* intervals for ε_C , and were also calculated using flat *priors* $\text{Exp}(0.001)$ for μ and σ . Table VI summarizes our results. As expected, the *posterior* mean for ε_C was closer to 1 when the *prior* 95 per cent credible interval for ε_C was narrower; but the *posterior* 95 per cent credible interval were only slightly changed. The Bayes factors were not overly dependent on the *prior* distributions, although some systematic trends were observed: for s ranging from 0.1 to 1, $\log_{10} \text{BF}(\varepsilon_C > 1)$ and $\log_{10} \text{BF}(\mu_C > \mu_A)$ increased by 0.18 and 0.10, respectively, while $\log_{10} \text{BF}(\alpha_C > \alpha_A)$ and $\log_{10} \text{BF}(\beta_C > \beta_A)$ decreased by 0.15 and 0.34, respectively. The *priors* for the duration parameters also had little effect on the Bayes factors and did not change our conclusions.

7. DISCUSSION

The statistical model described in this article allowed estimation of the main epidemiological characteristics of influenza transmission in households. Estimation in this model was challenging because a large part of the infectious process was not observed, but it proved possible to estimate simultaneously the instantaneous risks of infection and the duration of the infectious period.

The follow-up of influenza-like symptoms in households with laboratory confirmed index cases was used to estimate the parameters for influenza infectiveness. A crucial point of the study was therefore to define, in the observation level, a link between symptoms and infectiveness. Little information exists on this issue. It may be that the intensity of symptoms mirrors infectiveness, and this would allow efficient use of all the available data. However, the validity of such assumptions is questionable, and prompted us to use a *minimal* observation level, resting on the well-grounded hypothesis of short incubation time for influenza. While this choice used only the first day of symptoms in the likelihood, it enabled to estimate transmission parameters, showing that much information was contained in the sequence of appearance of cases.

We used longitudinal data, which clearly provided information on the risk of intra household transmission, but also on the community risk. Indeed, in 30 per cent of the households, the index case was the only case in the household during the 15 days of follow-up. Therefore, discarding approximately 4 days during which this infectious subject was present in the household, and put the other members at risk, there was up to 11 days during which households members were exposed to (and escaped) community risk only. The effect of censoring after 15 days was likely to be small for intra-household parameters since, with an estimated average duration of infectious period of 3.8 days, in good agreement with previously reported values [18–20], 15-days follow-up appeared enough to observe a chain of transmission in one household. This was confirmed by simulations: we found that the bias was small for all parameters, but could reach 20 per cent for the community risk. However, data simulated with a longer follow-up in each household (1 month), showed that this bias could be greatly reduced.

Dealing with the absence of uninfected households is a critical issue, which has previously been overcome by the use of conditional likelihood [1]. Here, the model was conditioned on the state of the household at the beginning of the follow-up (1 index case, the other members being susceptible), so that the probability for the initial state of the household does not appear in the likelihood. We checked by simulating complete epidemics, i.e. with uninfected households, that a change in the proportion of uninfected households had no effect on the bias in the community risk (results not shown).

Another concern is that we based the identification of cases on clinical diagnoses, assuming that subjects with clinical influenza were infectious, while those without clinical influenza were not. This may not be as informative as serologically or virologically confirmed diagnoses. In our data set, subjects were sampled from households with a virologically confirmed index case. The large percentage (75 per cent) of contact cases that became ill shortly (in the 5 days) after the index case suggests that the latter should have been responsible for a large portion of the contact cases. This conclusion agreed with simulations of the transmission model, where the index case was found responsible for at least 73 per cent of the contact cases. Therefore, the prevalence of influenza infection must have been high in the

households, and it is therefore expected that the positive predictive value and the negative predictive value of clinical influenza were large. Challenges on volunteers have shown that most infected individuals shedding the virus were clinically ill [14, 16], which suggests it is reasonable to neglect asymptomatic carriage.

The two main hypotheses concerning the model for the natural history of influenza were that (i) the natural history of influenza can be modeled with an SIR model, (ii) influenza incubation period is between 1 and 3 days. The first hypothesis implied that a subject became infectious as soon as he was infected. This assumption seemed acceptable considering challenges on volunteers [14–17] in which influenza virus was detected shortly after virus inoculation (24h). The second hypothesis was consistent with the literature [13–17].

All estimates were obtained using vague *prior* information, except for the duration of the infectious period where the *prior* was somewhat informative. However, we checked that using the flat prior $\text{Exp}(0.001)$ for the parameters μ and σ , the *posterior* distributions were not modified. As expected for influenza, the duration of the infectious period appeared to be shorter than the duration of symptoms: while the median duration for symptoms was 7 days with [4, 11] as inter-quartile interval, the median duration of the infectious period was 3.5 with [2.4, 4.9] as inter-quartile interval. Therefore, the model adjusted for a short infectious period, even if no constraint had been put on the end of the infectious period, relative to the end of the symptoms.

We found that to obtain a good fit to the data, the risk of infection had to be roughly proportional to the density of infectives rather than to their count. This fact disagrees with the hypothesis adopted in Longini *et al.* [1, 2]. Rather, it points to heterogeneity in transmission depending on the size of the household, whereby the number of effective contacts between an infective and a susceptible would decrease with the size of the household.

Since the risk of infection was dependent on the size of the household, the values of the SAR could not be directly compared to others obtained with the assumption of no such dependence. For comparison purposes, we however estimated our model with $\eta = 0$, and calculated the $\text{SAR} = 0.26[0.22, 0.29]$ and $\text{CPI} = 0.09[0.05, 0.14]$. These estimates were in line with those reported from a follow-up study of A/H3N2 influenza in Seattle [2], where the SAR was estimated at $0.20[0.12–0.29]$, and $\text{CPI} = 0.26[0.20–0.31]$. However, since the CPI was calculated over the whole length of the epidemic, rather than on 15 days as in here, the CPI was larger than our estimate.

Longini *et al.* estimation procedure gave excellent fit for household data [2] and was found quite robust for parameter values preset within appropriate limits for influenza [24]. Unexpectedly, fitting this model to our data yielded a poor fit ($\chi^2 = 20.1$, 10 df, $P = 0.028$). This lack of fit may result first from the manifest dependence of the risk of infection on the density of infectives rather than on their count, as reported above; it may also stem from the fact that our data were truncated, with households included only if there was at least one case. While the method of Longini *et al.* may be applied in such a setting [1], it is not known how this affects the validity of the results.

Bayes factors were used to investigate several epidemiological hypotheses. The simulations suggest that Bayes factors allowed us to discriminate between hypotheses $H1–H5$ since they were approximately centered on 0 under $S0$ (on the \log_{10} scale), and they were sensitive to a modest change in parameters as studied in scenarios $S1–S4$. Our analysis confirmed that children have a specific role in influenza transmission [4]. There was substantial evidence that children are more susceptible than adults. One could have expected a stronger evidence

for this hypothesis, because children are reportedly more susceptible to influenza infection: before an epidemic, low antibody levels were found in 84 per cent preschool children, as opposed to 49 per cent and 53 per cent in adults [25]. Unfortunately, there were not enough very young subjects in the present data set to examine this issue in more detail. We found substantial evidence that children have a larger community risk than adults. The risk ratio $\alpha_C/\alpha_A \approx \text{CPI}_C/\text{CPI}_A \approx 1.5$ was similar to previous estimations [25]. There was poor evidence that children are infectious for a shorter duration than adults. It would have been interesting to detail the dependence of the duration of infectious period on age as it was shown that very young children with influenza shed virus for up to 10 days compared with between 3 and 5 days reported in adults and adolescents [26]. Here again, the small number of very young subjects prevented this analysis. The main specificity of children in this study was that they are more infectious than adults. There was indeed decisive evidence in favour of this hypothesis and the household risk was found to be twice as large for an infectious child than for an infectious adult. This evidence should be taken into account in evaluating prevention strategies.

APPENDIX A

We present here the procedure to simulate an epidemic with parameters $(\alpha_C, \alpha_A, \beta_C, \beta_A, \mu_C, \mu_A, \varepsilon_C, \sigma)$ in a household of size n , with age distribution $\{a_1^*, \dots, a_n^*\}$.

The age a_1 of the first infected is randomly drawn in $\{a_1^*, \dots, a_n^*\}$. The date of the first infection v_1 is drawn uniformly in $[0, 1]$ and the corresponding duration of infectious period $\psi_1 - v_1$ is drawn from the gamma distribution with mean and standard deviation (μ_C, σ_C) if $a_1 \leq 15$, and (μ_A, σ_A) if $a_1 > 15$.

At time v_k corresponding to the k th infection in the household, we denote $\{a_1, \dots, a_k\}$ the ages of the infectives, $\{\psi_1, \dots, \psi_k\}$ the dates of end of infectious period, and $\{a_1^*, \dots, a_{n-k}^*\}$ the ages of the susceptibles. We define $S_C^k = \sum_{i=1}^{n-k} \mathbf{1}\{a_i^* \leq 15\}$ the number of susceptible children and $S_A^k = \sum_{i=1}^{n-k} \mathbf{1}\{a_i^* > 15\}$ the number of susceptible adults at time v_k . For $t > v_k$, we define $I_C^k(t) = \sum_{i=1}^k \mathbf{1}\{t < \psi_i; a_i \leq 15\}$ the number of infectious children, and $I_A^k(t) = \sum_{i=1}^k \mathbf{1}\{t < \psi_i; a_i > 15\}$ the number of infectious adults at time t among the first k infectives of the household.

Before the $(k + 1)$ th infection occurs, the instantaneous risk that a susceptible child is infected at time $t > v_k$ is

$$\lambda_C(t) = \alpha_C + \frac{\varepsilon_C}{n} \{\beta_C I_C^k(t) + \beta_A I_A^k(t)\}$$

and the instantaneous risk that a susceptible adult is infected at time $t > v_k$ is

$$\lambda_A(t) = \alpha_A + \frac{1}{n} \{\beta_C I_C^k(t) + \beta_A I_A^k(t)\}$$

Given v_k , the cumulative distribution function of the date of the $(k + 1)$ th infection v_{k+1} , defined in $[v_k, +\infty[$ is

$$F_{k+1}(t) = 1 - \exp \left\{ - \int_{v_k}^t (S_C^k \lambda_C(u) + S_A^k \lambda_A(u)) du \right\}$$

v_{k+1} is defined by solving $F_{k+1}(v_{k+1}) = r$, where r is a uniform deviate in $[0, 1]$.

Given v_{k+1} , the age a_{k+1} of the $(k + 1)$ th infective is randomly drawn among the ages of the susceptible children with probability $P_{k+1} = S_C^k \lambda_C(v_{k+1}) / (S_C^k \lambda_C(v_{k+1}) + S_A^k \lambda_A(v_{k+1}))$, and among the ages of the susceptible adults with probability $1 - P_{k+1}$.

The corresponding duration of infectious period $\psi_{k+1} - v_{k+1}$ is drawn from the gamma distribution with mean and standard deviation (μ_C, σ_C) if $a_{k+1} \leq 15$, and (μ_A, σ_A) if $a_{k+1} > 15$. The process is iterated until there is no more susceptible.

For a subject i infected at time v_i , the duration d_i of the incubation period is drawn from the uniform distribution $U[0, 3]$, i.e. the day Z_i^* when clinical influenza starts is $|v_i + d_i|$. In order to standardize the results, we assume that (i) clinical influenza is first observed in the household at day 0 so that we consider $Z_i = Z_i^* - \min_{j \in \{1, \dots, n\}} Z_j^*$; (ii) the observation period is 15 days after the first clinical influenza occurrence so that we observe Z_i only if $Z_i < 15$.

REFERENCES

1. Longini IM, Koopman JS. Household and community transmission parameters from final distributions of infections in households. *Biometrics* 1982; **38**(1):115–126.
2. Longini IM, Koopman JS, Monto AS, Fox JP. Estimating household and community transmission parameters for influenza. *American Journal of Epidemiology* 1982; **115**(5):736–751.
3. Haber M, Longini IM, Cotsonis GA. Models for the statistical analysis of infectious disease data. *Biometrics* 1988; **44**(1):163–173.
4. Addy CL, Longini IM, Haber M. A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* 1991; **47**(3):961–974.
5. Carrat F, Sahler C, Leruez M, Roger S, Freymuth F, Le Gales C, Bungener M, Housset B, Nicolas M, Rouzioux C. Influenza burden of illness: estimates from a national prospective survey of household contacts in France. *Archives of Internal Medicine* 2002; **162**(16):1842–1848.
6. Becker NG. *Analysis of Infectious Disease Data*. Chapman & Hall: London, 1989.
7. Gibson GJ. Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *Applied Statistics* 1997; **46**(2):215–233.
8. Gibson GJ, Renshaw E. Estimating parameters in stochastic compartmental models using Markov chain methods. *IMA Journal of Mathematics Applied in Medicine and Biology* 1998; **15**:19–40.
9. Auranen K, Arjas E, Leino T, Takala AK. Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. *Journal of the American Statistical Association* 2000; **95**(452):1044–1053.
10. Rampey Jr AH, Longini Jr IM, Haber M, Monto AS. A discrete-time model for the statistical analysis of infectious disease incidence data. *Biometrics* 1992; **48**(1):117–128.
11. Bailey NTJ. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin: London, 1975.
12. Anderson RM, May RM. *Infectious Diseases of Humans, Dynamics and Control*. Oxford University Press: New York, 1991; 15.
13. Nicholson KG. Human Influenza. In *Textbook of Influenza*, Nicholson KG, Webster RG, Hay AJ (eds). Blackwell Science: London, 1998; 219–264.
14. Fritz RS, Hayden FG, Calfee DP, Cass LMR, Peng AW, Alvord WG, Strober W, Straus SE. Nasal cytokine and chemokine responses in experimental influenza A virus infection: results of a placebo-controlled trial of intravenous Zanamivir treatment. *The Journal of Infectious Diseases* 1999; **180**:586–593.
15. Reuman PD, Bernstein DI, Keefer MC, Young EC, Sherwood JR, Schiff GM. Efficacy and safety of low dosage amantadine hydrochloride as prophylaxis for influenza A. *Antiviral Research* 1989; **11**(1):27–40.
16. Hayden FG, Treanor JJ, Betts RF, Lobo M, Esinhart JD, Hussey EK. Safety and efficacy of the neuraminidase inhibitor GG167 in experimental human influenza. *Journal of the American Medical Association* 1996; **275**(4):295–299.
17. Hayden FG, Treanor JJ, Fritz RS, Lobo M, Betts RF, Miller M, Kinnersley N, Mills RG, Ward P, Straus SE. Use of the oral neuraminidase inhibitor oseltamivir in experimental human influenza: randomized controlled trials for prevention and treatment. *Journal of the American Medical Association* 1999; **282**(13):1240–1246.
18. Treanor JJ, Hayden FG. Volunteer challenge studies. In *Textbook of Influenza*, Nicholson KG, Webster RG, Hay AJ (eds). Blackwell Science: London, 1998; 524–526.
19. Rvachev LA, Longini IM. A mathematical model for the global spread of influenza. *Mathematical Biosciences* 1985; **75**:3–22.
20. Flahault A, Letrait S, Blin P, Hazout S, Ménarés J, Valleron AJ. Modelling the 1985 influenza epidemic in France. *Statistics in Medicine* 1988; **7**:1147–1155.

21. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, 1996.
22. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**:457–511.
23. Robert C. *The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation* (2nd edn). Springer: New York, 2001.
24. Longini IM, Seaholm SK, Ackerman E, Koopman JS, Monto AS. Simulation studies of influenza epidemics: assessment of parameter estimation and sensitivity. *International Journal of Epidemiology* 1984; **13**(4): 496–501.
25. Longini IM, Koopman JS, Haber M, Cotsonis GA. Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. *American Journal of Epidemiology* 1988; **128**(4):845–859.
26. Welliver R, Monto AS, Carewicz O, Schatteman E, Hassman M, Hedrick J, Jackson HC, Huson L, Ward P, Oxford JS. Effectiveness of oseltamivir in preventing influenza in household contacts: a randomized controlled trial. *Journal of the American Medical Association* 2001; **285**(6):748–754.
27. Carrat F, Tachet A, Rouzioux C, Housset B, Valleron AJ. Evaluation of clinical case definitions of influenza: detailed investigation of patients during the 1995–1996 epidemic in France. *Clinical Infectious Diseases* 1999; **28**:283–290.