



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Analysis of household data on influenza epidemic with Bayesian hierarchical model



C.Y. Hsu^{a,b}, A.M.F. Yen^c, L.S. Chen^c, H.H. Chen^{b,*}

^a Department of Emergency Medicine, Taipei City Hospital, Zhongxing Branch, Taipei, Taiwan

^b Division of Biostatistics, College of Public Health, National Taiwan University, Taipei, Taiwan

^c School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

ARTICLE INFO

Article history:

Received 17 May 2013

Revised 14 November 2014

Accepted 22 November 2014

Available online 4 December 2014

Keywords:

Bayesian hierarchical model

Infectious disease

Influenza

Chain binomial model

ABSTRACT

Data used for modelling the household transmission of infectious diseases, such as influenza, have inherent multilevel structures and correlated property, which make the widely used conventional infectious disease transmission models (including the Greenwood model and the Reed–Frost model) not directly applicable within the context of a household (due to the crowded domestic condition or socioeconomic status of the household). Thus, at the household level, the effects resulting from individual-level factors, such as vaccination, may be confounded or modified in some way. We proposed the Bayesian hierarchical random-effects (random intercepts and random slopes) model under the context of generalised linear model to capture heterogeneity and variation on the individual, generation, and household levels. It was applied to empirical surveillance data on the influenza epidemic in Taiwan. The parameters of interest were estimated by using the Markov chain Monte Carlo method in conjunction with the Bayesian directed acyclic graphical models. Comparisons between models were made using the deviance information criterion. Based on the result of the random-slope Bayesian hierarchical method under the context of the Reed–Frost transmission model, the regression coefficient regarding the protective effect of vaccination varied statistically significantly from household to household. The result of such a heterogeneity was robust to the use of different prior distributions (including non-informative, sceptical, and enthusiastic ones). By integrating out the uncertainty of the parameters of the posterior distribution, the predictive distribution was computed to forecast the number of influenza cases allowing for random-household effect.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The statistical analysis of infectious disease data with probability models can not only elucidate how pathogens spread across susceptible population but also identify factors accounting for the transmission of infectious diseases. Specifically, it is of great interest to quantify the force of an infectious agent's spread and the associated factors. However, it should be noted that data on the transmission of pathogens during an epidemic period have inherent multilevel structures, from the individual, through the household, then finally to the community, using the transmission of influenza as an example. Moreover, when it comes to the evolution of the same epidemic wave infectious process across generations can also be treated as a separate level. Correlations between generations and across hierarchical levels make intractable the elucidation of the spread of infectious disease

based on conventional probability models. Because the transmission of pathogens at the community level is so different from that at the household level, it is therefore interesting in the study of some infectious diseases, such as influenza, to model the transmission mode based on household data before applying the model to community data.

Becker's chain binomial model is a seminal instrument for elucidating the nature of infectious disease transmission using the correlated property between successive observations during the propagation of epidemics within households [1]. Stemming from the chain binomial model, Becker's proposed linear logistic models allow for the escape probability that depends on relevant covariates of interest (such as the number of exposed infectives and generations) [1]. Alternative methods of modelling the transmission of diseases using household data with final size distributions have been proposed [2–4]. Using such an approach, Longini and Koopman estimated the escape probability of influenza within a household [2]. House and Keeling applied deterministic dynamic models to demonstrate the effects of strategies targeted at an individual or household while considering the heterogeneity of different household sizes [5,6]. By applying the

* Corresponding author at: Room 533, No. 17, Hsu-Chow Road, Taipei 100, Taiwan. Tel.: +886 2 33668033; fax: +886 2 33668042.

E-mail address: chenlin@ntu.edu.tw (H. Chen).

methods analogous to the branching process, Ball et al. extended the susceptible-infected-recovered (SIR) model to incorporate two types of mixing patterns (within-group and between-group transmission) and demonstrated that a control strategy should take into account certain heterogeneities, such as households [3,7,8]. The two separate model parameters pertaining to infection rates underpinning 2-level mixing epidemic model have been estimated by a data augmentation method using the entire at-risk population [9,10] proposed by Demiris and O'Neill or sample data proposed by O'Neill [11].

When data are analysed without considering temporal order, these methods provide a solution that considers possible latent epidemic chains. Although the use of a final size distribution alleviates the difficulty of identifying the temporal order of the onset of the disease, it may lose information on the generation time effect when data on the evolution of the epidemic wave are available. Namely, these models dispense with the need for classifying household cases as the same episode and explicitly model within and between household infections. Cauchemez et al. proposed a Bayesian approach to elucidate the transmission of influenza within households using prospective follow up data on the disease onset in households but accurate temporal information on the onset of the disease is not available [12]. Note that the majority of these models put emphasis on the threshold of epidemic based on the estimation of reproductive number with and without considering the containment of infection control, particularly universal vaccination. While these statistical models under the context of the SIR model are well developed for quantifying infectious process so as to evaluate the effectiveness of different strategies of infection control by using the estimated reproductive number the perspective of their application is at population (macro) level. Their models, although still workable we suppose by using the multi-type extended models for example [9–11], have been barely extended to model the heterogeneity of multilevel levels making allowance for correlation within each level.

To this end, we extended Becker's framework to accommodate data on such a hierarchical structure. Levels of susceptible individual, generation, and household were considered as a unified analytical framework under the context of generalised linear random-effect models [13]. Such a typical example can be noted in the O'Neill et al. study [14] that extended the flexibility of the Becker model by using the random household effect models in conjunction with the MCMC method to model the escaped probabilities at community and household level each of which was specified by a statistical distribution to account for the heterogeneities of the levels. Although their approach has advanced in the Becker model it still leaves room to be desired and may have several extensions. First, the mechanisms of disease transmission such as the Greenwood model [15] and the Reed–Frost model [16] originally proposed by the original Becker's framework can be incorporated to make the model biologically plausible. Second, the heterogeneity of escape probability across households (socioeconomic status) and individual attributes (such as age and sex) can be also taken into account by using the random-intercept model. Third, allowing for regression coefficients to vary with different levels can be accommodated by using the random-slope model. The classical example of this situation is that the effect of vaccination on protecting individuals from disease vulnerability should also consider contextual factors, such as the heterogeneity of socioeconomic status across households, which have been confirmed as significant factors in previous studies [5,6,17–20]. In spite of the attractiveness of these random-effect models, it should be noted that considering the effect and variation across different levels increases the number of parameters and makes estimation of parameters intractable from the viewpoint of statistical computation. The most serious problem is the problem of identifiability between parameters already noted in the O'Neill study [14]. To tackle these statistical technique problems, we resorted to the application of Bayesian underpinning together with the two transmission models.

Our major objective is to propose a framework incorporating the aforementioned multilevel data structure inherited from the epidemic data on influenza using Bayesian generalised linear models to facilitate hypothesis testing. We constructed hierarchical models with the following levels: subject, generation, and household. Correlations and contextual effects were modelled as the parameters of random intercepts and random slopes in the model. The MCMC method, in conjunction with the Bayesian directed acyclic graphical (DAG) model, was utilised to derive posterior distributions for the parameters of interest. The rest of this article is organised as follows. In Section 2, we demonstrate the application of our model to surveillance data on an influenza epidemic. Section 3 presents the Becker model and the proposed Bayesian hierarchical models based on the Greenwood and Reed–Frost transmission models. Following the Becker method, the logistic regression form was used to accommodate the effects of factors at an individual level. Section 4 describes the procedures of parameter estimation and model building, including the elicitation of prior distributions, Gibbs sampling, and model selection. The results of the analyses, as well as the predictive distribution and estimated effects of vaccination by the elicitation of informative priors to incorporate prior knowledge on the effects of vaccination, are presented in Section 5, and we conclude with a discussion.

2. Data source and motivation

Our intention in constructing Bayesian hierarchical models was motivated by the multilevel data structure of the observations of the influenza epidemic within household information. Between July 2001 and March 2002, Taiwanese surveillance data were collected to cover the yearly period of seasonal influenza [21]. Data on whether subjects developed influenza-like symptoms were derived using the National Health Insurance database. Cases of influenza were identified according to clinical diagnosis by a physician in the form of International Classification of Disease, 9th Revision, clinical modification (ICD-9-CM) codes 487 (influenza), 487.0 (influenza with pneumonia), 487.1 (influenza with respiratory manifestations), and 487.8 (influenza with other manifestations). Information on the characteristics of the subjects, such as age (divided into three groups: less than or equal to 6 years, ≥ 65 years, and 7–64 years), sex, vaccination status, and date of diagnosis were also collected. Household information was identified by linking the current empirical data with the nationwide household registration database of Taiwan, thus providing a unique identification number for each household. Following the definition of the Becker model [1], the first case in a household, identified by the date of diagnosis, was classified as the introductory case. Subsequent cases within the same household that had intervals between successive dates of diagnosis ≤ 7 days were classified as cases of subsequent generations, and the other household members free from influenza moved to the next generation [22,23]. Cases within the household diagnosed on the same day were classified as cases within the same generation. If the interval between successive dates of diagnosis was >7 days, the case was classified as the next episode (wave) of an epidemic. Data used for further analysis were confined to subjects with household sizes ranging from two to five with at least one case of influenza. The subjects used for estimating the parameters of the models were derived from a random sample of 10% of the households providing data.

Considering the data collected on the epidemic of influenza, at least two nested levels exist: the household level and the individual level nested within the household level. In addition to the characteristics of each susceptible subject, such as age, sex, and vaccination status, the environmental contexts of each household, including environmental crowding, the social network, and socioeconomic status, are thought to affect the spread of disease. This effect can result in variations of the baseline risk of contracting influenza between households, as has been observed in previous studies and was further

modelled using a random household effect [1,24,25]. Aside from the correlations between subjects within each household, heterogeneity exists between generations as an epidemic evolves. Such a generation effect has been previously modelled in the Becker model [1] and was also observed in the work of Fraser et al. [26] during the H1N1 strain epidemic. Variations within the joint effects of factors at the individual level and at the household level represent the heterogeneities inherited from these subpopulations. In addition to variation in the baseline risk of contracting influenza across households, the effect of the influenza vaccine, which should be evaluated after the implementation of a risk-based vaccination project including the voluntary uptake of vaccine, may also vary from household to household due to the interaction between the environmental contexts and the effects of vaccination at the individual level. This variation should also be taken into account.

These correlations within a household, inherited from the data, result in a violation of common assumptions, such as independence and homogeneity between observations, and lead to the phenomenon of overdispersion in the construction of probability models. Transmission probabilities change across households and generations and vary between individuals with different characteristics. The nested relationships between levels are also observed. As a result, a multi-level structure occurs in our data on the influenza epidemic, and the source of this variation arises from levels that should be jointly evaluated in combination with relevant factors at an individual level. As mentioned above, we are motivated by this intuition to construct Bayesian hierarchical models to accommodate these correlational structures while preserving the mechanism of transmission (such as the Greenwood model and the Reed–Frost model) by which disease spreads within households to analyse the collected data.

3. Model specification

3.1. Becker's linear logistic model

Under the framework of the chain binomial model, the effect of each risk factor on the escape probability can be assessed by standard methods of analysing binary outcomes. Becker [1] proposed a linear logistic model allowing for the escape probability, depending on the relevant covariates of interest. Let t act as a non-negative integer denoting the generation of the evolution of infectives within a household, with introductory cases belonging to generation zero. Let Y_{tj} indicate whether a susceptible subject escaped from disease after being exposed to j infectives in the generation $t - 1$, acting as a Bernoulli random variable with the parameter q_{tj} . The Bernoulli parameter represents the escape probability indicated by number of exposed infectives j and generation t , and the probability of being vulnerable to disease, p_{tj} , is $1 - q_{tj}$. The logistic form $\log\{(m_{tj}q_{tj})/[m_{tj}(1 - q_{tj})]\} = \alpha_t + \beta_t \times j$ was adopted in Becker's generalised linear model [1], with m_{tj} denoting the number of susceptibles exposed to j infectives in generation $t - 1$. The effects of generations and the number of infectives on the escape probability were thus assessed by the model. The Greenwood model [15], which assumes the risk of being infected for susceptible subjects is independent of the number of infectives (saturated infection), is equivalent to reducing α_t to α and specifying β_t as zero. On the other hand, the Reed–Frost model [16] asserts that the event of escaping infection when exposed simultaneously to two infectives is equivalent to escaping infection when exposed to one infective in each of two successive occasions and depicts the transmission mechanism of close contact, which is equivalent to reducing β_t to β . Applying Becker's generalised linear model to our data, we are able to incorporate factors at an individual level alongside the generation level and the exposed number of infectives in the model to evaluate their combined effects on the risk of susceptibility to influenza. This gives

$$\text{logit}(p_{htk}) = \alpha_t + \beta_0 \times j_{ht} + \beta' \mathbf{X}_{htk}, \tag{1}$$

where p_{htk} denotes the probability of becoming an influenza case of subject k in generation t and household h , which is the complement of the escape probability q_{htk} in the Becker model. Thus, the random variable is Y_{htk} , which represents the realisation of empirical data on the epidemic of influenza for a susceptible subject k of generation t and household h escaping from influenza. The number of infectives j_{ht} is considered a generation-level covariate that is nested within a household, and \mathbf{X}_{htk} denotes the vector of covariates of the individual k , nested in generation t and household h , after including age, gender, and vaccination status into our data. β is the vector of the corresponding regression coefficients. Because of the mild nature of influenza, we assume the homogeneous generation effect (β_0) rather than the time-varying property (β_t) on the number of infectives. This is Becker's original linear logistic model, considering the generation effect with a Reed–Frost underpinning. Assuming a constant baseline probability of cases' becoming influenza across generations, Becker's linear logistic model without the generation effect based on the Reed–Frost transmission model can be written as follows:

$$\text{logit}(p_{htk}) = \alpha + \beta_0 \times j_{ht} + \beta' \mathbf{X}_{htk}.$$

Becker's generalised linear model can be further modified using the Greenwood transmission model as the mechanism of the spread of influenza, which gives

$$\text{logit}(p_{htk}) = \alpha_t + \beta' \mathbf{X}_{htk}$$

for the Greenwood model with the generation effect and

$$\text{logit}(p_{htk}) = \alpha + \beta' \mathbf{X}_{htk}$$

for the Greenwood model without the generation effect.

3.2. Bayesian acyclic graphical model for multilevel data

To estimate the parameters with the hierarchical random effects of households and effects of individual-level covariates, a methodology developed for the analysis of DAG models was utilised. To address the multilevel data structure, several statistical methods have been proposed, including the individual-specific random effects model [27,28] and the population-based average method [29]. The former, using the Bayesian DAG model, has been widely used in the literature [30–32]. Details of the DAG model have been described by Spiegelhalter et al. [33].

In terms of multilevel property, we have three levels in the data: the individual level, generation level, and household level. According to the process by which the data gathered, the levels were assumed to be nested within one another: individuals within generation and generations within households. As in Becker's linear logistic model, we expressed the model using the indexes k, t , and h for the individual, generation and household levels, respectively. The baseline risk for a susceptible subject's being vulnerable to influenza may vary from household to household due to contextual effects at the household level. This argument gives

$$\text{logit}(p_{htk}) = \alpha_h + \beta_0 \times j_{ht} + \beta' \mathbf{X}_{htk}, \tag{2}$$

where α_h captures the variation of a case's baseline risk of developing influenza across households, forming a random-household effect in the model.

3.2.1. Hierarchical models based on the Greenwood model

We first extend Becker's linear logistic model based on the Greenwood model to incorporate the variation of contextual effects across households. The model specified in (2) is simplified to give

$$\begin{aligned} \text{logit}(p_{htk}) &= \alpha_h + \beta' \mathbf{X}_{htk} \\ \alpha_h &= \alpha_{00} + \alpha_{0h}, \\ \alpha_{0h} &\sim N(0, \sigma_{\alpha_h}^2) \end{aligned} \tag{3}$$

where the common intercept α_{00} represents the average baseline risk of being vulnerable to influenza. The normal distributed random variable α_{0h} with a mean of zero and the variance, denoted as $\sigma_{\alpha_h}^2$, captures the heterogeneity at a household level (a random intercept with a cluster effect on household). The parameter corresponding to the effect of the number of infectives is zero ($\beta_0 = 0$) as the Greenwood model is adopted.

The effect of each generation on the probability of experiencing influenza cases is also of great interest and can be a separated baseline risk as α_t from the variation across households (α_h). The model specified in (3) can be extended to give

$$\begin{aligned} \text{logit}(p_{htk}) &= \alpha_h + \alpha_t + \beta' \mathbf{X}_{htk} \\ \alpha_h &= \alpha_{00} + \alpha_{0h} \\ \alpha_{0h} &\sim N(0, \sigma_{\alpha_h}^2). \end{aligned} \tag{4}$$

As in Becker's linear logistic model, α_t represents the effect of each generation on the baseline probability of turning into influenza cases. The normal variate α_{0h} models the variation across the household level (a random-intercept model with a cluster effect on household, considering the generation effect).

Aside from the variation in the baseline probability of producing influenza cases, the contextual effect may interact with the effect of vaccination in reducing susceptibility to influenza, especially when the uptake of vaccination is voluntary. To model the variation of vaccination's effect across households, the model in (3) is modified accordingly to give

$$\begin{aligned} \text{logit}(p_{htk}) &= \alpha + \beta_h \times V_{htk} + \beta' \mathbf{X}_{htk} \\ \beta_h &= \beta_{00} + \beta_{0h}, \\ \beta_{0h} &\sim N(0, \sigma_{\beta_h}^2) \end{aligned} \tag{5}$$

where V_{htk} is the vaccination status of subject k of generation t and household h , and β_h represents the effect of vaccination on the probability of being susceptible to influenza with common value β_{00} and normal variate β_{0h} , allowing the value to change from household to household. Parameter $\sigma_{\beta_h}^2$ captures the extent of this variation (a random slope with a cluster effect on household). The model specified in (5) can be extended with a similar concept as (4) to produce a model with a generation effect and a random slope reflecting the interaction between the effect of vaccination and environmental context of a household, based on the Greenwood transmission model (a random-slope model with a cluster effect on household, considering the generation effect):

$$\begin{aligned} \text{logit}(p_{htk}) &= \alpha + \alpha_t + \beta_h \times V_{htk} + \beta' \mathbf{X}_{htk} \\ \beta_h &= \beta_{00} + \beta_{0h}. \\ \beta_{0h} &\sim N(0, \sigma_{\beta_h}^2) \end{aligned} \tag{6}$$

3.2.2. Hierarchical models based on the Reed–Frost model

Hierarchical models based on the Reed–Frost model as the mechanism of disease transmission, incorporating variation across households (a random intercept with a cluster effect on household), are constructed by specifying the model in (2) as follows:

$$\begin{aligned} \text{logit}(p_{htk}) &= \alpha_h + \beta_0 \times j_{ht} + \beta' \mathbf{X}_{htk} \\ \alpha_h &= \alpha_{00} + \alpha_{0h}, \\ \alpha_{0h} &\sim N(0, \sigma_{\alpha_h}^2) \end{aligned} \tag{7}$$

where the zero-mean normal variate α_{0h} with variance denoted as $\sigma_{\alpha_h}^2$ captures the variation of the contextual effect on the risk of contracting influenza across households (a random intercept with a cluster effect on household). Parameter β_0 represents the effect of number of infectives on the probability of contracting influenza, which corresponding to the Reed–Frost transmission model. Similarly, based on

the Reed–Frost model, a hierarchical model that accounts for variations in the effect of vaccination across households is given by modifying (2) as follows:

$$\begin{aligned} \text{logit}(p_{htk}) &= \alpha + \beta_0 \times j_{ht} + \beta_h \times V_{htk} + \beta' \mathbf{X}_{htk} \\ \beta_h &= \beta_{00} + \beta_{0h} \\ \beta_{0h} &\sim N(0, \sigma_{\beta_h}^2). \end{aligned} \tag{8}$$

As in the hierarchical models based on the Greenwood model, the effect of each generation on the risk of turning into influenza cases can be separated from the baseline risk by adding the α_t term to (7) and (8) to give corresponding hierarchical models that consider the generation effect, based on the Reed–Frost model.

An example of the DAG representation of the proposed Bayesian hierarchical models based on the Reed–Frost model is given in Fig. 1. Models incorporating random intercepts (black eclipse nodes) and random slopes (grey eclipse nodes) are both depicted. Considering the hierarchical model based on the Reed–Frost underpinning, given the parameter $q[h, t, k]$, the random variable $Y[h, t, k]$ represents whether a susceptible k of household h and generation t escaped from influenza. The random variable follows a Bernoulli distribution, and the probability of being vulnerable to influenza for the subject $p[h, t, k]$ is the complement of $q[h, t, k]$, which is in turn linked with the covariates at an individual level (age[h,t,k], sex[h,t,k], and vaccination[h,t,k]), with the intercept changing for each household (α_h in the model with random intercept). The household-specific intercept (α_h) is then further decomposed into the average value of the baseline risk and the variation across households (α_{00} and α_{0h} , black eclipse nodes), modelling the heterogeneity of the baseline risk at the household level. Considering the model with random slope, the effect of vaccination for each subject allows for a household-specific value (β_h), which is further decomposed into the average effect and variation across households (β_{00} and β_{0h} , grey eclipse nodes), corresponding to the interaction between the contextual effect at the household level and the effect of vaccination at the individual level. The number of infectives who can expose influenza to the susceptible of household h and generation t denoted by j_{ht} is also included in the systematic component to accommodate the Reed–Frost transmission model. Following Becker's linear logistic model, a logit link is adopted. Systematic components and the variations that arose from the level of household are specified in (7) and (8) for the model with random intercepts (black eclipse nodes) and the one with random slopes (grey eclipse nodes), respectively, as depicted in Fig. 1.

4. Estimation and model selection

4.1. Likelihood function

Following the Becker model, random variable Y_{htk} can be modelled using Bernoulli distribution:

$$\begin{aligned} Y_{htk} &= \begin{cases} 1 & \text{if escaped from disease} \\ 0 & \text{otherwise} \end{cases} \\ Y_{htk} | q_{htk} &\sim \text{Bernoulli}(q_{htk}). \\ p_{htk} &= 1 - q_{htk} \\ 0 &\leq q_{htk} \leq 1 \end{aligned}$$

The overall likelihood function for the empirical data is thus given by

$$L = \prod_{h=1}^{\text{households}} \prod_{t=1}^{\text{generations}} \prod_{k=1}^{\text{subjects}} q_{htk}^{Y_{htk}} (1 - q_{htk})^{1 - Y_{htk}}.$$

4.2. Assignment of the prior distribution

Non-informative priors $N(0, 10^6)$ were used for individual-level parameters β as well as for the parameter of the effect of the

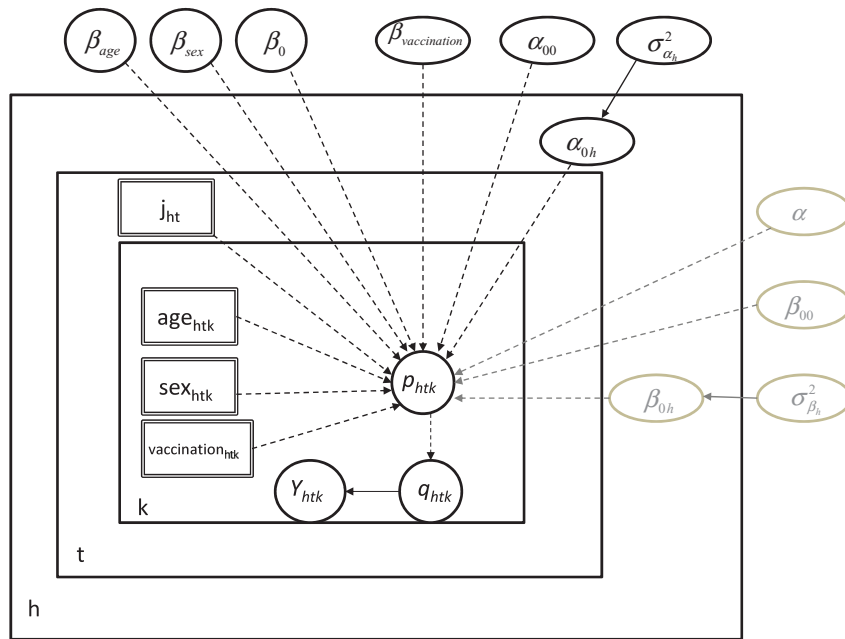


Fig. 1. Acyclic graphical model for Bayesian hierarchical models based on the Reed–Frost model.

Model with random intercepts (black eclipse nodes)

$$\begin{aligned}
 Y_{htk} | q_{htk} &\sim \text{Bernoulli}(q_{htk}) \\
 p_{htk} &= 1 - q_{htk} \\
 \text{logit}(p_{htk}) &= \alpha_h + \beta_0 \times j_{ht} + \beta_{\text{age}} \times \text{age}_{htk} + \beta_{\text{sex}} \times \text{sex}_{htk} + \beta_{\text{vaccination}} \times \text{vaccination}_{htk} \\
 \alpha_h &= \alpha_{00} + \alpha_{0h} \\
 \alpha_{0h} &\sim N(0, \sigma_{\alpha_h}^2)
 \end{aligned}$$

Model with random slopes (grey eclipse nodes)

$$\begin{aligned}
 Y_{htk} | q_{htk} &\sim \text{Bernoulli}(q_{htk}) \\
 p_{htk} &= 1 - q_{htk} \\
 \text{logit}(p_{htk}) &= \alpha + \beta_0 \times j_{ht} + \beta_{\text{age}} \times \text{age}_{htk} + \beta_{\text{sex}} \times \text{sex}_{htk} + \beta_h \times \text{vaccination}_{htk} \\
 \beta_h &= \beta_{00} + \beta_{0h} \\
 \beta_{0h} &\sim N(0, \sigma_{\beta_h}^2).
 \end{aligned}$$

number of infectives (β_0) on the risk of being vulnerable to influenza. In models containing such random-household effects, non-informative priors $N(0, 10^6)$ were assigned for the regression coefficients α_{00} , representing the average risk in random-intercept models, and β_{00} , representing the average effect of vaccination in random-slope models. In hierarchical models, the random variables capture the variation across households, α_{0h} and β_{0h} , which were modelled using the normal distributions $N(0, \sigma_{\alpha_h}^2)$ and $N(0, \sigma_{\beta_h}^2)$, respectively. The non-informative hyperpriors with uniform distributions $U(0, 10^3)$ were assigned to the square roots of the variances σ_{α_h} and σ_{β_h} . In hierarchical models considering generation effects, the non-informative priors $N(0, 10^6)$ were assigned to the regression coefficients α_t . In addition to non-informative priors, informative priors, including sceptical and enthusiastic ones, on the regression coefficient of vaccination (β_{00}) were elicited to evaluate their effects on the posterior distribution. By reviewing the evidence provided in the Cochrane Library, the effect of reducing the risk of influenza-like illness (ILI) by vaccination was estimated at 10%–35% [34,35]. Based on this prior knowledge and assuming an exchangeable property, we were able to assign the sceptical prior to β_{00} according to the logarithm of one minus the quantity of risk reduction as $N(0, 7 \times 10^{-3})$, with the enthusiastic prior as $N(-0.27, 7 \times 10^{-3})$. Details of the derivation of the informative priors based on the results of a literature review are provided in Appendix A in the supplementary material.

4.3. Gibbs sampling, model selection, and the predictive distribution

A Gibbs sampler was used to derive samples of a stationary distribution by which inferences on posterior distributions were made. The initial values for the regression coefficients, including the parameters of individual-level factors (β), the parameter of number of infectives (β_0), and the parameters of the average values of the random effects at the household level (α_{00} and β_{00}), were set as zero. In the hierarchical models with random intercepts and random slopes, the initial values of the hyperpriors of σ_{α_h} and σ_{β_h} were set at one. Full conditional distributions based on the DAG of models were used to update the process of sampling. All of the estimates of parameters to derive summary statistics of posterior distributions were computed by 5000 iterations carried out with a thinning interval of three after a burn-in period of 5000 iterations. Such a MCMC simulation yielded a total of 1667 updated posterior samples. Estimates based on the posterior distributions of parameters were derived using the MCMC method, which was carried out using WinBUGS [36]. Examples of the Bayesian hierarchical model with random intercepts based on the Greenwood model and based on the Reed–Frost model using WinBUGS based on DAG are given in Appendix B in the supplementary material. Posterior distributions for the parameters of other proposed models were evaluated similarly.

Comparisons between models were guided by the deviance information criterion (DIC). Using the posterior estimates of the proposed

Table 1
Characteristics of the study subjects^a.

	Study subjects		Sampled households of sizes 2–5 ^b							
	Flu	Non-flu	2		3		4		5	
	Flu	Non-flu	Flu	Non-flu	Flu	Non-flu	Flu	Non-flu	Flu	Non-flu
Number	6,616,738 (29.5)	15,834,203 (70.5)	92,027 (52.4)	80,397 (46.6)	117,036 (38.9)	183,798 (61.1)	144,933 (31.8)	311,283 (68.2)	100,291 (26.8)	273,429 (73.2)
Gender (male)	3,153,296 (47.7)	8,304,946 (52.5)	41,929 (45.6)	42,214 (52.5)	56,569 (48.3)	96,351 (52.4)	71,425 (49.3)	165,382 (53.1)	47,485 (47.4)	139,981 (51.2)
Vaccine	272,073 (4.1)	656,150 (4.1)	8948 (9.7)	7854 (9.8)	4131 (3.5)	8164 (4.4)	2115 (1.5)	7091 (2.3)	1396 (1.4)	7231 (2.6)
Age (SD) ^c	30.6 (21.9)	35.8 (19.6)	40.1 (23.1)	44.3 (19.2)	29.0 (21.7)	36.1 (19.0)	25.7 (19.1)	33.1 (17.7)	25.6 (18.9)	32.8 (18.3)

^a Data are presented as numbers (%) of subjects, unless otherwise stated.

^b Subjects of sampled households with more than one influenza case.

^c Average (standard deviation) of age.

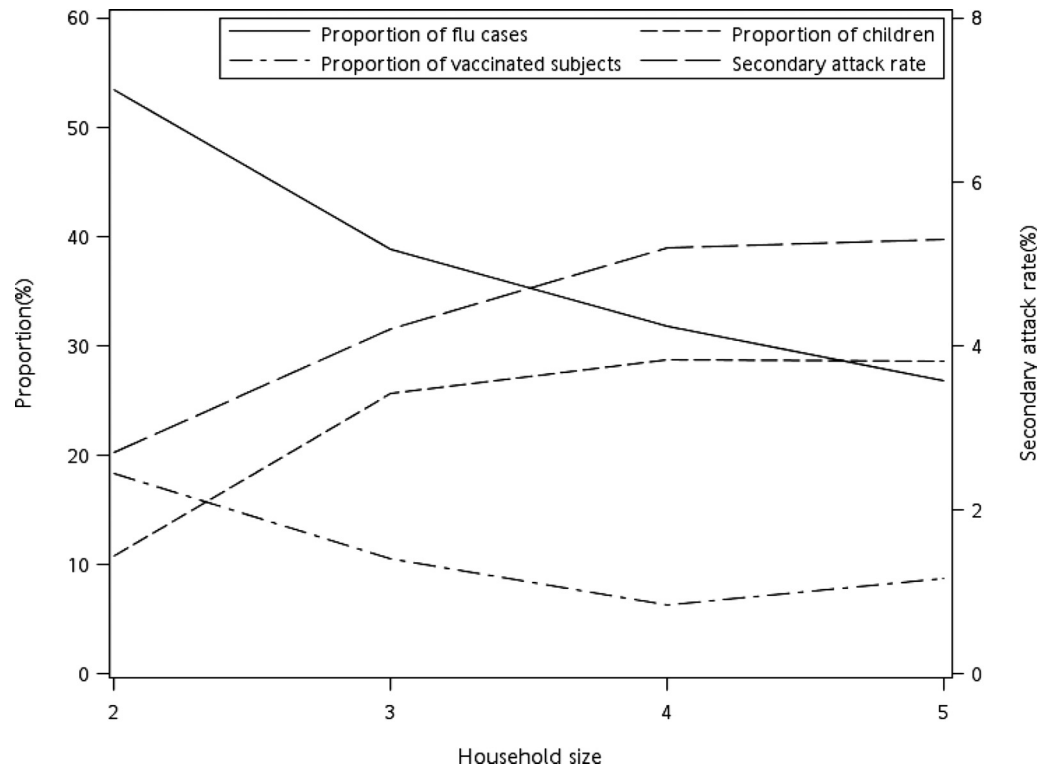


Fig. 2. Heterogeneity across households in terms of the proportion of influenza cases (solid line), proportion of vaccinated subjects (dotted dashed line), proportion of children (short dashed line), and secondary attack rate (long dashed line).

Bayesian hierarchical models, we were able to give a predictive distribution. The prediction was applied to another set of randomly sampled data consisting of 456,224 subjects and 27,782 influenza cases derived from 168,060 households. Based on the results of model comparisons, the predictive distribution of the number of influenza cases according to the household sizes and vaccination statuses using models with the Reed–Frost underpinning were evaluated to demonstrate the consequences of including contextual effects in the model.

5. Results

5.1. Descriptive findings

During the study period, 6,616,738 flu cases were observed among the 22,450,941 people living in Taiwan. From these cases, a total of 13,039,481 subjects derived from 3,752,035 households of sizes ranging from two to five with at least one influenza case were collected.

Using the sampling fraction of 10%, 1,303,194 subjects derived from 375,288 households were enrolled to construct the hierarchical models. The mean age of the sampled subjects was 32.9 (standard deviation: 19.8) years, and the proportions of males and of those vaccinated were 50.8% and 3.6%, respectively. Table 1 shows the characteristics of the study population and the sampled subjects, stratified by household size. The average proportion of household members diagnosed as influenza cases decreased from 52.4% in households of size two to 26.8% in households of size five, with corresponding secondary attack rates of 2.7%, 4.2%, 5.2%, and 5.3% in households of sizes two, three, four, and five, respectively. The data used for analysis were confined to households with at least one case, including the introductory and secondary cases. Fig. 2 shows these differences across the sizes of households alone, including the proportions of children (≤ 6 years of age) and vaccinated subjects. This figure represents the heterogeneity between households that may affect the probability of influenza transmission within households.

Table 2
Estimated results of Becker’s linear logistic models (estimates (95% credible interval)).

		Greenwood model		Greenwood model with a generation effect		Reed–Frost model		Reed–Frost model with a generation effect	
Intercept		–2.77	(–2.78, –2.76)	–2.81	(–2.82, –2.79)	–2.78	(–2.79, –2.77)	–2.82	(–2.83, –2.81)
Number of infectives		–	–	–	–	0.28	(0.25, 0.32)	0.29	(0.25, 0.33)
Age	≤6	1.50	(1.48, 1.53)	1.50	(1.48, 1.53)	1.50	(1.48, 1.53)	1.50	(1.48, 1.52)
	≥65	–0.34	(–0.40, –0.29)	–0.33	(–0.38, –0.28)	–0.34	(–0.39, –0.29)	–0.33	(–0.38, –0.27)
Sex (male)		–0.22	(–0.23, –0.20)	–0.22	(–0.23, –0.20)	–0.21	(–0.23, –0.20)	–0.22	(–0.23, –0.20)
Vaccination		–0.04	(–0.12, 0.03)	–0.04	(–0.11, 0.03)	–0.04	(–0.12, 0.03)	–0.04	(–0.11, 0.04)
Generation effect	Second	–	–	0.29	(0.27, 0.32)	–	–	0.30	(0.27, 0.33)
	Third	–	–	0.42	(0.34, 0.50)	–	–	0.43	(0.35, 0.50)
	Fourth	–	–	0.80	(0.46, 1.09)	–	–	0.81	(0.46, 1.11)

Table 3
Estimated results of Bayesian hierarchical models based on the Greenwood model (estimates (95% credible interval)).

		Random intercept				Random slope (vaccination status)			
		Greenwood model		Greenwood model with a generation effect		Greenwood model		Greenwood model with a generation effect	
Intercept	α_{00}	–2.84	(–2.85, –2.82)	–2.81	(–2.82, –2.79)	–2.77	(–2.78, –2.76)	–2.81	(–2.82, –2.79)
	σ_{α_h}	0.35	(0.31, 0.38)	0.02	(0.00, 0.07)	–	–	–	–
Age	≤6	1.52	(1.50, 1.55)	1.50	(1.48, 1.53)	1.50	(1.48, 1.52)	1.50	(1.48, 1.53)
	≥65	–0.34	(–0.39, –0.29)	–0.33	(–0.38, –0.28)	–0.35	(–0.40, –0.30)	–0.34	(–0.39, –0.28)
Sex (male)		–0.22	(–0.23, –0.20)	–0.22	(–0.23, –0.20)	–0.22	(–0.23, –0.20)	–0.22	(–0.23, –0.20)
Vaccination	β_{00}	–0.04	(–0.12, 0.03)	–0.04	(–0.11, 0.04)	–0.58	(–1.00, –0.23)	–0.64	(–1.02, –0.32)
	σ_{β_h}	–	–	–	–	1.11	(0.66, 1.52)	1.18	(0.81, 1.54)
Generation effect	Second	–	–	0.29	(0.27, 0.32)	–	–	0.30	(0.27, 0.32)
	Third	–	–	0.42	(0.34, 0.49)	–	–	0.42	(0.34, 0.50)
	Fourth	–	–	0.79	(0.47, 1.09)	–	–	0.80	(0.47, 1.11)

5.2. Becker’s linear logistic model with and without a generation effect

The estimated results of Becker’s linear logistic model based on the Greenwood model and the Reed–Frost model, with and without the generation effect, are shown in Table 2. The Greenwood model and Reed–Frost model revealed similar trends when covariates at the individual level were considered. Compared with young adults, the probability of being vulnerable to influenza was higher for children (≤6 years of age) and lower for the elderly (>65 years). Males had a lower risk of contracting influenza. The estimated result of the regression coefficient for the number of infectives in the Reed–Frost model was 0.28 (95% credible interval (CI): 0.25–0.32), which manifests the effect of increasing the risk of being vulnerable to influenza, with an approximately 30% increase in the odds per increment of infectives (Table 2, third column). The estimated effects of vaccination on the risk of turning into influenza cases were –0.04 (95% CI: –0.12 to 0.03), using both the Greenwood model and the Reed–Frost model, suggesting the protective effect of vaccination, although both of the ranges included zero (Table 2, first and third columns). The trend of increasing the probability of being vulnerable to influenza by generations was observed in both models incorporating a generation effects (Table 2, second and fourth columns).

5.3. Bayesian hierarchical models based on the Greenwood model

Table 3 shows the estimated results of applying the Bayesian hierarchical models containing random intercepts and random slopes with a cluster effect on households (σ_{α_h} and σ_{β_h} , respectively) based on the Greenwood underpinning. In the random-intercept Bayesian hierarchical model, the estimated variation of the household (σ_{α_h}) was 0.35 (95% CI: 0.31–0.38), which was approximately 12% of the estimated result of average baseline risk, α_{00} (Table 3, first column). Considering the random-intercept model with a generation effect, the variation at the household level decreased to 0.02 (95% CI: 0.00–0.07), which was 0.7% of the estimated result of α_{00} (Table 3, second

column). The estimated effect of vaccination (β_{00}) using the model with random intercepts was similar to that of Becker’s linear logistic model and also included a 95% CI containing zero. Considering the random-slope model, the estimated average effect of vaccination (β_{00}) was –0.58 (95% CI: –1.00 to –0.23), which corresponded to the protective effect of vaccination and an odds ratio of 0.56 (95% CI: 0.37–0.80) (Table 3, third column). A similar protective effect of vaccination was observed in the random-slope model with a generation effect (Table 3, fourth column). The variation of the household level for the effect of vaccination (σ_{β_h}) was 1.11 (95% CI: 0.66–1.52) which was almost twice the estimated value of the common effect, β_{00} . A similar extent of the variation of vaccination effects across households was still observed after including a generation effect in the model (Table 3, third and fourth columns).

These results suggest that the variation in the baseline risk of being vulnerable to influenza was reduced substantially after a generation effect was considered. However, the vaccination effect statistically significantly varied from household to household, as was still shown even after including the generation effect into the model. In the models considering the interaction between the effects of vaccination and household, reflecting the contextual effect of the environment, the estimated protective effect of vaccination was enhanced, and the 95% CI departed from zero. To show the differences among Becker’s linear logistic model, the Bayesian hierarchical model with random intercepts, and the latter model with random slopes, we illustrated these three models using estimated results based on the Greenwood model with a generation effect by simulating the logit predicted values of the vaccinated and unvaccinated subjects of ten households and by controlling for covariates at the subject level, considering ages between 6 and 65 years and females. Fig. 3 shows the results of the simulation of four generations, as arranged from top to bottom. The predicted values of 10 households are represented by 10 line patterns. The slope of each line represents the estimated effect of vaccinating the specific household estimated by the three models. For Becker’s linear logistic model based on the Greenwood model,

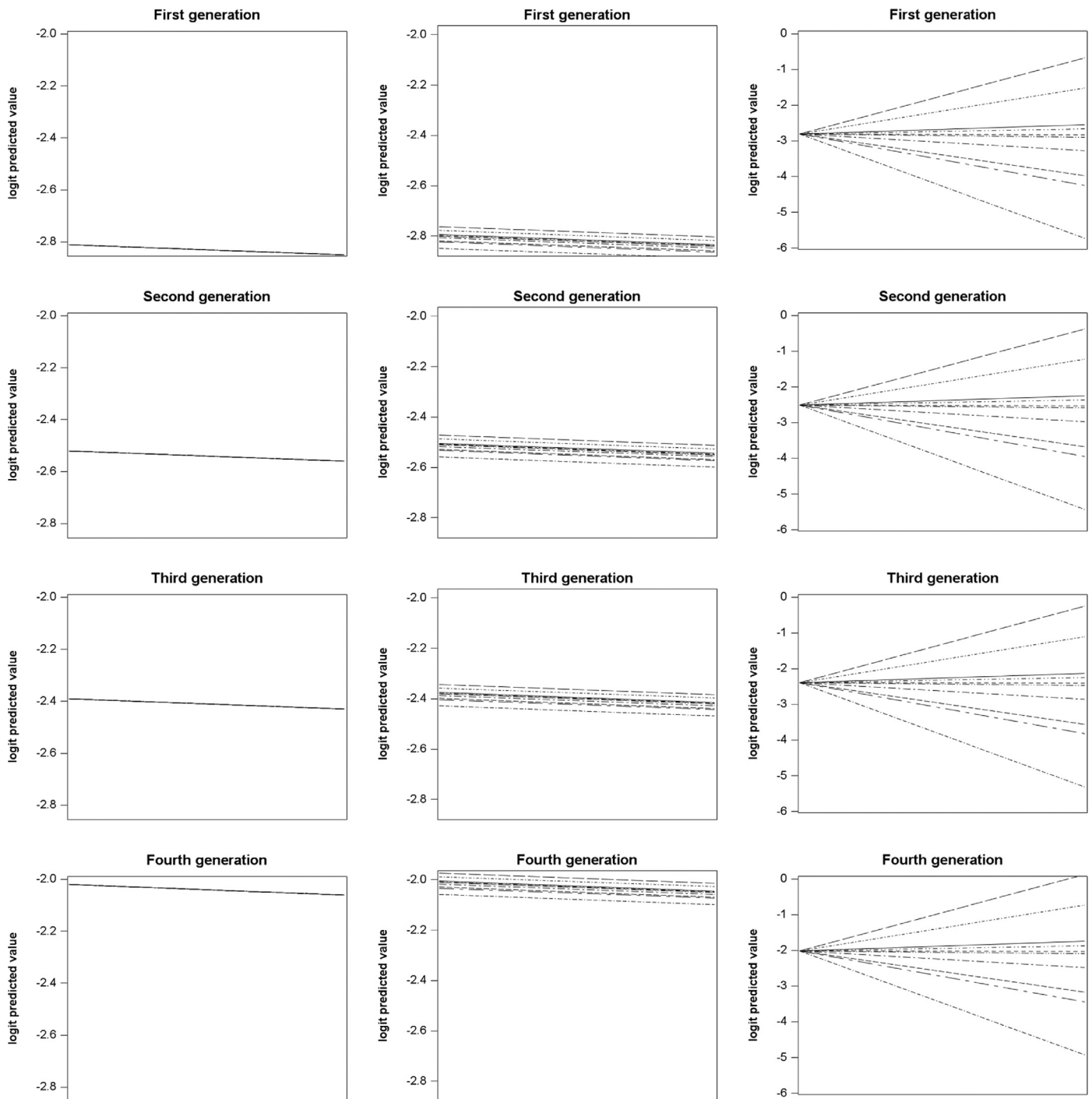


Fig. 3. Graphical illustration of the difference between Becker's linear logistic model and the Bayesian hierarchical models using the Greenwood model with a generation effect for the effect of vaccination on the logit predicted value (vertical axis). The figures from top to bottom represent the results from the first to fourth generations. Simulated results of ten households are represented by the ten different line patterns in each figure.

there was no variation between households, and 10 lines overlapped (Fig. 3, first column). The intercepts of the logit-predicted values increased from -2.81 in the first generation (Fig. 3, first column and first row) to -2.02 in the fourth generation (Fig. 3, first column and fourth row), showing that the risk of being vulnerable to influenza increased as the epidemic evolved. In the random-intercept model based on the Greenwood model (Fig. 3, second column), the trend of increasing the risk of being vulnerable to influenza with successive generations was preserved, and the spread of the logit-predicted values of the ten simulated household are shown as parallel lines

due to the α_{0h} term in the model specified in (4), which represents the heterogeneity in the baseline risk of being vulnerable to influenza across households and the contextual effect. Regarding the Greenwood model with a random slope (Fig. 3, third column), the variation between households shows the effect on the slope (β_{0h} , model specified in (8)), representing the interaction between the effect of vaccination and the contextual effect. Moreover, it is notable that the heterogeneity in the slope is much greater than that of the intercept after separating the generation effect from the baseline risk.

Table 4
Estimated results of Bayesian hierarchical models based on the Reed–Frost model (estimates (95% credible interval)).

		Random intercept				Random slope (vaccination status)			
		Reed–Frost model		Reed–Frost model with a generation effect		Reed–Frost model		Reed–Frost model with a generation effect	
Intercept	α_{00}	-2.84	(-2.86, -2.83)	-2.82	(-2.83, -2.81)	-2.78	(-2.80, -2.77)	-2.82	(-2.83, -2.81)
	σ_{α_h}	0.34	(0.30, 0.37)	0.02	(0.00, 0.07)	-	-	-	-
Number of infectives		0.28	(0.24, 0.32)	0.29	(0.25, 0.33)	0.28	(0.25, 0.33)	0.29	(0.26, 0.33)
Age	≤ 6	1.52	(1.50, 1.54)	1.50	(1.48, 1.52)	1.50	(1.48, 1.53)	1.50	(1.48, 1.52)
	≥ 65	-0.34	(-0.39, -0.28)	-0.33	(-0.38, -0.28)	-0.34	(-0.39, -0.29)	-0.33	(-0.39, -0.28)
Sex (male)		-0.22	(-0.23, -0.20)	-0.22	(-0.23, -0.20)	-0.22	(-0.23, -0.20)	-0.22	(-0.23, -0.20)
Vaccination	β_{00}	-0.04	(-0.11, 0.04)	-0.04	(-0.11, 0.03)	-0.63	(-0.95, -0.29)	-0.69	(-1.15, -0.29)
Generation effect	σ_{β_h}	-	-	-	-	1.14	(0.76, 1.49)	1.23	(0.77, 1.66)
	Second	-	-	0.30	(0.27, 0.32)	-	-	0.30	(0.27, 0.32)
	Third	-	-	0.43	(0.35, 0.50)	-	-	0.43	(0.35, 0.51)
	Fourth	-	-	0.81	(0.48, 1.12)	-	-	0.81	(0.50, 1.11)

Table 5
Comparison of DIC for models.

Model	DIC	Dbar	pD
Becker's linear logistic models			
Greenwood model	406,496	406,491	5
Greenwood model with a generation effect	405,931	405,923	8
Reed–Frost model	406,298	406,292	6
Reed–Frost model with a generation effect	405,721	405,712	9
Bayesian hierarchical model, random intercept			
Greenwood model	406,168	399,677	6490
Greenwood model with a fixed generation effect	405,897	405,879	19
Reed–Frost model	406,029	399,823	6206
Reed–Frost model with a fixed generation effect	405,641	405,672	-30
Bayesian Hierarchical model, random slope (vaccination status)			
Greenwood model	406,178	405,117	1061
Greenwood model with a fixed generation effect	405,686	404,401	1285
Reed–Frost model	406,020	404,839	1181
Reed–Frost model with a fixed generation effect	405,300	404,083	1217

5.4. Bayesian hierarchical models with the Reed–Frost model

Modelling the mechanism of influenza transmission within households as resulting from close contact, we extended Becker's linear logistic model based on the Reed–Frost model to build hierarchical models, with the estimated results shown in Table 4. The effect of the number of infectives was similar in how it described increased risk of contracting influenza to the corresponding Becker linear logistic model. Similar to the Bayesian hierarchical models based on the Greenwood model, the reduction in the estimated variation across households (α_{0h}) after including the generation effect in the model was observed to decrease from 0.34 to 0.02 (Table 4, first and second columns). Using the random-intercept model, the estimated protective effect of vaccination was -0.04 (β_{00}) with the 95% CI including zero (Table 4, first and second columns). This quantity moved away from zero using the random-slope model, incorporating the interaction between the effect of vaccination and the household-level context and was estimated as -0.63 to -0.69 , with a corresponding odds ratio of 0.53 to 0.50 (Table 4, third and fourth columns).

5.5. Comparison between models

Table 5 presents the DIC values of a variety of models. As far as Becker's linear logistic models are concerned, the Reed–Frost model had lower DIC values than the Greenwood model (viz., a difference of 198). Table 5 also shows that making allowances for the generation effect would reduce the DIC value by 565 based on the Greenwood model. The corresponding reduction of DIC value by 577 was noted for the Reed–Frost model, suggesting the necessity of modelling the generation effect.

Compared with their fixed-effect counterparts in Becker's linear logistic models, which allowed for variation across households (a random effect), the reduced DIC values figured substantially in those Bayesian hierarchical models with random intercepts and with random slopes. The random-intercept model based on the Greenwood model resulted in the reduction of the DIC value by 328 compared with the Greenwood model of the fixed-effect model. A corresponding 269 reduction in the DIC values was noted for the Reed–Frost model. With respect to the Bayesian hierarchical models with random slopes, the model based on the Greenwood model reduced the DIC values by 318 compared with the fixed-effect model based on the Greenwood model. A corresponding reduction of the DIC value by 278 was noted in the comparison between models based on the Reed–Frost model. However, the reduction in the DIC value by including the random intercept parameter in the models already containing the generation effect was modest (34 for the Greenwood model and 80 for the Reed–Frost model). This finding suggests the significant contribution of generation effect. The comparisons between the fixed-effect models and the random-effect models demonstrate the significance of including random-effect parameters (viz., σ_{α_h} and σ_{β_h}) to improve the conventional probability model by taking the heterogeneity between households into account when tackling the phenomenon of overdispersion. Considering a random slope for generations in the model reduced the DIC value further (211 for the Greenwood model and 341 for the Reed–Frost model) compared with the model with random intercepts. This result suggests that the heterogeneity accounting for the overdispersion in modelling the process of the influenza epidemic lies mainly in the variation of the effects of vaccination across households, after including the generation effect in the model.

To summarise, the Bayesian hierarchical model with random slopes based on the Reed–Frost model with a generation effect incorporating random variation across households into the effect of vaccination (random slope) is the model with the lowest DIC value (405,300, Table 5). The next-best model is the random-intercept Bayesian hierarchical model with a generation effect based on the Reed–Frost model (405,641, Table 5). Details on simulated posterior samples of parameters, including autocorrelation plots, correlations between parameters, and effective sample sizes, are provided in Appendix C.

5.6. Estimated results by the elicitation of informative priors

Applying the Bayesian approach offers the advantage of introducing prior distributions and thus has the potential to incorporate knowledge on the spread of influenza based on previous studies or observations of epidemics into the model. Informative priors, including sceptical and enthusiastic ones on the protective effect of vaccination, were incorporated. Combined with the weight of the evidence derived from the collected data using the proposed Bayesian

Table 6

Estimated results of Bayesian hierarchical models with random slopes using sceptical priors and enthusiastic priors on the effect of vaccination (estimate (95% credible interval)).

		Greenwood model with a generation effect				Reed–Frost model with a generation effect			
		Enthusiastic prior DIC: 405,871		Sceptical prior DIC: 405,525		Enthusiastic prior DIC: 405,657		Sceptical prior DIC: 405,340	
Intercept	α_{00}	-2.81	(-2.82, -2.79)	-2.81	(-2.82, -2.79)	-2.82	(-2.83, -2.81)	-2.82	(-2.83, -2.81)
Number of infectives		-	-	-	-	0.29	(0.25, 0.33)	0.29	(0.25, 0.33)
Age	≤ 6	1.50	(1.48, 1.53)	1.50	(1.48, 1.52)	1.50	(1.48, 1.52)	1.50	(1.48, 1.52)
	≥ 65	-0.34	(-0.39, -0.29)	-0.34	(-0.40, -0.29)	-0.33	(-0.39, -0.28)	-0.34	(-0.39, -0.29)
Sex (male)		-0.22	(-0.23, -0.20)	-0.22	(-0.23, -0.20)	-0.22	(-0.23, -0.20)	-0.21	(-0.23, -0.20)
Vaccination	β_{00}	-0.32	(-0.48, -0.18)	-0.10	(-0.23, 0.01)	-0.32	(-0.45, -0.17)	-0.12	(-0.25, 0.00)
	α_{β_0}	0.81	(0.56, 1.03)	0.39	(0.03, 0.71)	0.80	(0.55, 1.00)	0.40	(0.07, 0.75)
Generation effect	Second	0.30	(0.27, 0.32)	0.30	(0.27, 0.32)	0.30	(0.27, 0.32)	0.30	(0.27, 0.33)
	Third	0.42	(0.34, 0.50)	0.42	(0.34, 0.50)	0.43	(0.35, 0.50)	0.43	(0.35, 0.50)
	Fourth	0.80	(0.47, 1.11)	0.79	(0.46, 1.12)	0.81	(0.50, 1.11)	0.81	(0.50, 1.11)

Enthusiastic prior: $\beta_{00} \sim N(-0.27, 7 \times 10^{-3})$; Sceptical prior: $\beta_{00} \sim N(0, 7 \times 10^{-3})$.

hierarchical models, we were able to obtain estimates on the protective effect, updated by data. Table 6 provides the estimated results and DIC values reflecting the use of these informative priors and random-slope models with generation effects based on the Greenwood model and the Reed–Frost model. Even if the sceptical prior showing prior belief in the effect of vaccination were to concentrate around the null, the estimated protective effect (β_{00}) would be -0.1 to -0.12 , corresponding to a risk reduction of 10%–12%, although the tail would extend across zero (95% CI: -0.23 to 0.01 in the Greenwood model and -0.25 to 0.00 in the Reed–Frost model). Using an enthusiastic prior showing a belief in risk reduction attributable to vaccination that concentrated at 25%, the estimated effect (β_{00}) updated by data was -0.32 for the Greenwood model and the Reed–Frost model. This figure corresponds to a risk reduction of approximately 30%, which lies between 25% (i.e., the prior belief) and 45% (i.e., the evidence provided by the data) and results from the average of the prior and the evidence derived from the data, as weighted by their strength. The comparison of the DIC between models with the two informative priors shows that the model with sceptical priors performs better.

5.7. Predictive distribution using models based on the Reed–Frost model

Fig. 4 shows the predicted number of influenza cases stratified by the sizes of households and vaccination statuses using the results of four models based on the Reed–Frost model. The figures from top to bottom in Fig. 4 represent the predicted number of cases using the fixed-effect model based on the Reed–Frost model, the fixed-effect model based on the Reed–Frost model with a generation effect, the random-intercept model based on the Reed–Frost model with a generation effect, and the random-slope model based on the Reed–Frost model with a generation effect. The predictions were applied to all subjects (Fig. 4, first column), to those not having been vaccinated (Fig. 4, second column), and to those who were vaccinated (Fig. 4, third column). Considering the results of the prediction as they were applied to the overall group and the unvaccinated group, the predicted number of influenza cases was close to the observed one, with a narrow 95% CI range for all models, except for the fixed-effect model based on the Reed–Frost model with a generation effect. Although the point estimates of the predicted number of cases were close to the observed values for the prediction using the fixed-effect model based on the Reed–Frost model with a generation effect (Fig. 4, second row), the ranges of the 95% CIs were wide. Considering the prediction applied to the unvaccinated group, the predicted results deviated from the observed numbers of cases substantially in almost all sizes of households for all models, owing to the low vaccination rates (4.1%, Table 1) and relative small sizes of samples compared with the unvaccinated group. Despite this deviation, the ranges of the 95% CIs for the fixed-effect model based on the Reed–Frost model with a generation effect and the random-slope model based on the Reed–Frost model

with a generation effect covered the observed number of cases in the vaccinated group.

6. Discussion

By using the Bayesian hierarchical model, our approach extended Becker's linear logistic model, which takes into account the correlations between observed cases in epidemic data across households while preserving the mechanisms of disease transmission within the model. The Bayesian hierarchical model has been used to cope with the heterogeneity found in longitudinal follow-up studies characterised by multilevel structures [30–32]. The flexible framework of the Bayesian hierarchical approach makes it feasible to model data with complex structures. O'Neill et al. proposed the model's use for partially observed household and temporal data by applying the MCMC method to incorporate the concept of random infectiousness and random susceptibility in the model [14]. Methods on data augmentation for the longitudinal household data of influenza outbreaks, such as the MCMC simulation, were also developed, although the problem of convergence when estimating parameters was noted and was coped with alternative MCMC algorithm [14]. This problem may result from issues of identifiability, mainly for the parameters pertinent to the two escaped probabilities from household and community. Following the Becker model [1], we extended the method of analysing infectious disease data with household structures by making use of the Bayesian hierarchical approach to tackle correlated features between observations. The proposed method provides a framework using random effects underpinned by a generalised linear model to address the dependent data for infectious diseases within a household. This is tantamount to the application of the Bayesian hierarchical model to address the issue of overdispersion, partly due to the multilevel structure and partly to the dependent property of the data.

There are several merits of our proposed methods. The first is the proposed Bayesian hierarchical models that are capable to dealing with correlated data with multilevel structure while incorporating the mechanisms of disease transmission including the Greenwood model and the Reed–Frost model. The second advantage is that in addition to evaluating the effect of individual attributes (such as age, sex, and vaccination status) on the force of disease transmission, the heterogeneity across households can be also taken into account and quantified by using the random-intercept and random-slope models. The third is the simplicity of applying proposed method by using WinBUGS to fit the model without the need for specialised software.

In addition to the correlated property and multilevel data structures, individual factors that were considered included sex, age, vaccination status, and generation. Male sex and old age groups showed lower probabilities of being vulnerable to influenza. As the epidemic evolved, the probability of contracting influenza was higher. The vaccination effect was not significant in Becker's linear logistic models

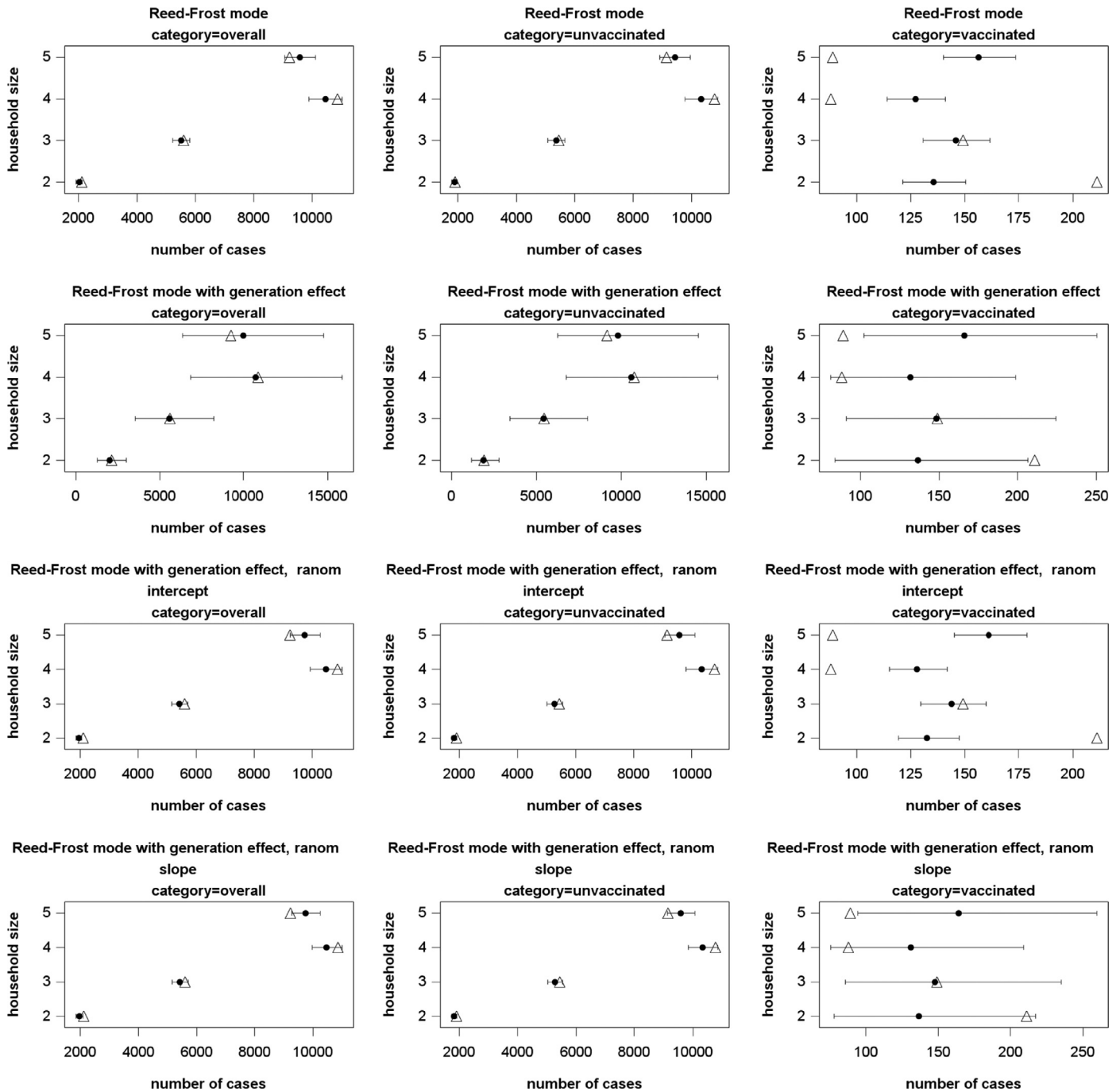


Fig. 4. Plots of the predicted (filled circle) and observed (hollow triangle) numbers of influenza cases (horizontal axis) and their corresponding 95% CIs against household sizes (vertical axis). The figures from top to bottom represent the results using the fixed-effect model based on the Reed–Frost model, the fixed-effect model based on the Reed effect model with a generation effect, the Bayesian hierarchical model with a random intercept based on the Reed–Frost model with a generation effect, and the Bayesian hierarchical model with random slope based on the Reed–Frost model with a generation effect.

or in the Bayesian hierarchical random-intercept models, whereas the protective effect of vaccination was statistically significant in the random-slope Bayesian hierarchical model. The difference in the risk of being vulnerable to influenza between males and females is very likely reflective of differential exposure to influenza infections. Regarding the differences observed among the age groups, the lower risk associated with the elder group and the higher risk associated with the younger group are consistent with the observations of Cauchemez et al. [24] and may be due to a history of previous exposure to the influenza virus that may confer a protective effect. The inverse relationship between the size of households and the probability of being

influenza cases listed in Table 1 was also observed in several studies [24, 37–39]. Although the trends of proportion of influenza cases by household size vary from study to study, possible factors such as the proportion of child [24], contact pattern, hygiene behaviour, ventilation status [40], and difference in social network pattern [39] implying the heterogeneity across households may contribute to the observed pattern.

The impression that the probability of contracting influenza differs from household to household is supported by variations within the environmental factors of each household, such as crowded domestic conditions, the proportion of the family consisting of children, the

contact patterns between household members, and the proportion of family members who have been vaccinated. A descriptive analysis of our data by the sizes of households also shows such variation (Fig. 2), prompting us to apply hierarchical models with random effects. Similar findings on variations across households were also observed by Cauchemez et al. [24] in their work on the pandemic of influenza. The work of House and Keeling also demonstrated that heterogeneity across households should be considered before implementing preventive strategies [5]. From the history of modelling infectious disease data, incorporating generation effects and variation across households are two approaches that have been found to be useful in improving the performance of models [1] due to their allowing for heterogeneity, which accounts for overdispersion. This approach is also supported by the fundamental theories of infectious diseases [41], which indicate that the delineation of the process of the spread of infectious disease between hosts will be fruitful only after the cluster effect arising from environmental contexts has been considered. Although the previous work on the methods of modelling the epidemic of influenza within households tackled the relevant difficulties, such as latent epidemic chains [2,3,14] and partially observed infectious processes [7–12,14], the approach of quantifying such heterogeneity has been barely addressed. The proposed Bayesian hierarchical framework on modelling infectious disease data also provides an approach to test the significance of random-effect parameters by comparing the DIC values between models. In our proposed models, both parameters of the random intercept (α_h) and the random slope (β_h) significantly improved the fitting of Becker's linear logistic model. After including a generation effect, which partially accounts for the baseline variation in the risk of contracting influenza, the improvement observed in the models with random intercepts diminished substantially in terms of DIC values. The comparison between the Bayesian hierarchical models with generation effects and random intercepts and those with random slopes shows that heterogeneity can be further addressed by considering the random-slope parameter, which reveals significant variation in the effect of vaccination across households. Significant improvements in model-fitting statistic in terms of DIC values were noted after incorporating a generation and random effects, as noted in Table 5. Among our Bayesian hierarchical models, the model with best performance in terms of DIC values was the random-slope model based on the Reed–Frost model with a generation effect, suggesting the existence of random variations on the effect of vaccination across households. Meanwhile, the probability of being vulnerable to influenza increased with respect to generation. The superior performance of the Reed–Frost model compared with the Greenwood model was also observed in Table 5 because the number of exposed infectives also increased the risk of being vulnerable to influenza during the spread of influenza within the households.

The contrast between random-effect models and the fixed-effect ones is presented in Table 5 showing the value of Dbar and pD for each model. In general, the DIC values of models with random-effect (random-intercept or random-slope) were reduced in comparison with the counterparts of the fixed-effect. This is entirely attributed to the lower value of Dbar of the random-effect model when compared with the counterpart of the fixed-effect. However, including the random effects in the model also led to the reduction in degree of freedom resulting from the complexity of the model. Such a penalty is represented by the higher value of pD of the random-effect model in comparison with the counterpart of the fixed-effect. The penalty of the elevated pD was offset by the gain of reducing Dbar. The net gains gave the support of random-effect models.

With respect to the comparisons made between the random-intercept models and the random-slope models, the estimated results based on the Bayesian hierarchical models with random intercepts showed some protective effect of vaccination, although the 95% CIs included zero. However, considering the interaction between the effects of vaccination and the contextual effects of households, the

estimated effect of vaccination using the Bayesian hierarchical models with random slopes departed from zero, with odds ratios of approximately 0.50 to 0.53. This paradox, when evaluated together with the low proportion of vaccination (4.1%, Table 1) and the voluntary nature of being vaccinated within the target population reveals the possibility that the significant protective effect of vaccination observed only in models with random slopes was due to the contextual effects of households, such as their socioeconomic statuses. Vaccination therefore represents a proxy for this context, which shows its effect in terms of interaction. In their work based on a survey of nationally representative samples in the United States for patients aged 18 and older, Galarce et al. [18] reported that receiving the influenza vaccine is associated with sociodemographic factors and other factors, such as urbanisation and influenza-related beliefs. Similar findings on the differences in receiving an influenza vaccination among people across socioeconomic strata have also been noted in Taiwan [19]. In addition to variation in receiving the influenza vaccination, the effect of the socioeconomic gradient on influenza was revealed in the work of Stone et al. [20]. Discrepancies between the vaccination rate and the risk of infection were also disclosed in the work of Gu and Sood [42]. These studies suggest possible links between socioeconomic factors, influenza vaccination, and the risk of being vulnerable to influenza. Using Bayesian hierarchical models with random slopes permits meaningful estimation of the vaccination effect across the level of households (σ_{β_h}), which quantifies this linkage in the form of an interaction. Additionally, the outcome used in our analysis was the clinical diagnosis of influenza as observed among individuals seeking medical consultations, which may also be subject to the socioeconomic status of each household.

In addition to elucidating the transmission of pathogens and associated factors, predicting the extent of an epidemic using the proposed model based on current data is another goal of modelling infectious disease. Based on the results of our prediction, the trends of the spreading infectious disease can be assessed, and a decision on the proper allocation of resources can be better made. For a prediction that more accurately reflects the progression of an epidemic, observable relevant factors, as well as unobserved variables, should be incorporated into the model. The Bayesian hierarchical approach has the advantage of incorporating unobserved variables, such as contextual effects and socioeconomic status, into the model in the form of random effects; therefore, it is able to provide a predictive distribution based on the current data, making use of the posterior estimates. The comparison of predictive distributions between the models presented in Fig. 4 shows the benefit of including random slopes into the model. The prediction using the Bayesian hierarchical model with random slopes based on the Reed–Frost model with a generation effect provides a precise result that is close to the observed value, with a narrow 95% CI range in the overall group and the unvaccinated group. Although the predicted result in the vaccinated group may be compromised, the ranges of the 95% CIs cover the observed values and can act as foundations upon which the extent of the epidemic may be assessed. The random-effect parameter on vaccination (β_h) provides the flexibility to account for heterogeneity across households. The low vaccination rate and the predominance of unvaccinated subjects in most of the influenza cases resulted in the predicted number of cases being largely influenced by the unvaccinated group, even after adjusting for age, sex, and heterogeneity across households.

The prior distributions elicited by the Bayesian approach provide a place to incorporate the results of previous observations and studies. When the non-informative prior was used, the posterior distributions reflected the weight of evidence based only on the data, whereas the elicitation of the informative prior provides an approach to evaluating the parameter of interest based on different prior beliefs that can be updated by the current data. The estimated results of the effect of the vaccination shown in Table 6, using the enthusiastic prior and the sceptical prior, illustrate this approach. Considering movements

both towards and away from the null and the effect of vaccinations, the posterior estimates provide results informed by the data. The weighting factors of the prior belief and the data are the inverses of their variances. In our case, the inverse of the variance of the prior distribution dominates the information derived from the data and shifts the estimated effect of vaccination towards the prior belief.

The MCMC method, combined with the DAG, provides an intuitive approach for constructing hierarchical models by representing the hierarchies between levels graphically, capturing the property of conditional independence, which further facilitates the derivation of full conditional posterior distributions used by the Gibbs sampler. Still, the MCMC method does not guarantee convergence or proper mixing during the sampling process. Reparameterisation is often required to improve performance [43]. Additionally, it is time consuming to generate sufficient posterior samples upon which inferences can be made; this process may be adversely affected by reparameterisation. This is especially true when applying hierarchical models to our empirical data and using households as cluster effects. Although the proposed method is flexible in accommodating the correlations across levels in data with hierarchical structures, identifiability was still unavoidable for certain parameter. Correlations of the posterior samples of parameters between the individual level and the generation level were observed for the effect of vaccination (β_{00}) and the random effect (σ_{β_h}), as illustrated in the scatter plots of models in Appendix C (Figs. C.2.3.1–C.2.3.4). To consider the correlation between β_{00} and σ_{β_h} , we applied the normal-inverse- χ^2 (NIC) method to model the joint effect of the two parameters [44]. The joint prior distribution $p(\beta, \sigma^2)$ is defined in terms of the marginal prior distribution for σ^2 and the conditional distribution of β given σ^2 , specifically, $p(\beta, \sigma^2) = p(\sigma^2)p(\beta|\sigma^2)$. The model specification based on the Greenwood model and the Reed–Frost model with a generation effect alongside the estimated results using sceptical prior are given in Appendix D in the supplementary material. Except for the standard deviation parameter (σ_{β_h}), estimated result for the effect of vaccination were robust to different prior distribution chosen (Table 6 and Table D in Appendix D).

The minimum effective sample sizes of 20–30 for the Bayesian hierarchical model with random slope based on the Greenwood model and the Reed–Frost model were due to the correlation between the effect of vaccination and its variation across households. Several iteration algorithms including longer thinning intervals, running multiple chains, and longer iterations were explored with the estimated results and effective sample size listed in Tables E.1–E.4 in Appendix E in the supplementary material. Although the effective sample size varied with these sampling algorithms, the estimated results were consistent. Moreover, the results based on three chains with different initial values proposed by Gelman et al. [43] also led to consistent estimates (see Figs. E.1 and E.2 in Appendix E). In addition to applying Gibbs sampling algorithm, we also explored the effective sample size, estimated results, and acceptance rate using rejection sampling (Tables E.5 and E.6 in Appendix E). The estimated results for parameters using rejecting sampling algorithm were consistent with those using Gibbs sampling. The results based on 5000 burn-in and 5000 iterations with a thinning interval of three were thus reported due to the consistency in the estimated results of these updating strategies. Based on these explorations, we think the estimation of some parameters of interest may be affected by high correlation but this issue can be ameliorated by longer simulation and validated by multiple chains and different sampling algorithms.

Using ILI as case definition makes it possible to utilise surveillance data to explore the transmission of disease and the effect of vaccination taking into account the relevant factors including individual level factors and heterogeneity across households. However, the nature of monitoring the disease status rather than the infectious status of the surveillance should be noted before making inference base on the results. The estimated secondary attack rate was lower than studies

based on prospective design [24,45]. Although the secondary attack rate within household ranged from <4% to >30% depends on the definition of case and area of study [46] and also determined by viral strain and household structure, the lower estimates of the secondary attack rate may associate with the selection of only subjects diagnosed as influenza rather than those with respiratory symptoms. Also, the identification of influenza cases relies on subjects with influenza seeking for medical consultation which may also contribute to the lower secondary attack rate. The use of clinical definition also introduces another source of variation, since virus such as *parainfluenza virus*, *adenovirus*, *coronavirus*, *respiratory syncytial virus*, *bacteria*, and other untypable virus may also cause ILI [47]. These underlying variations in the causative agent for ILI may contribute to the deviation between the predicted and observed number of cases presented in Fig. 4. Although nationwide laboratory surveillance network was unavailable until the year 2003 [48], recent studies showed the positive rate for influenza among subjects with ILI was around 22–27% and may reach 50% during epidemic [49,50]. They also indicated the parallel trend between reported ILI cases and the number of isolated influenza virus. These findings, together with the result of studies on the association between ILI and influenza infection [51,52] support the use of ILI as a proxy variable for the surveillance of number of isolated influenza virus.

What our proposed method is lacking of is to deal with the real epidemic data on within-group and between-group transmission of disease such as influenza, which is often modelled by two separate parameters with 2-level mixing epidemic models assuming independence to consider the heterogeneity of infection rates among different groups with good approximation while the population size is sufficiently large or the outbreak sufficiently small [7,8,14] or relaxing the assumption of independence with a data augmentation method [9–11]. Such a multi-type epidemic model considered local contact as well as global contact. Both are integrated into data augmentation scheme [9–11]. It is therefore worthwhile to make allowance for information on both local and global contact with our proposed Bayesian hierarchical model as considered in the O'Neill work [11] if such hierarchical data either with temporal outcome or final size outcome or a mixture of both at different population structures are available.

As far as the sampled data rather than the entire population used is concerned, it is very interesting to note that O'Neill has already proposed a very well generalised framework of data augmentation to deal with the sampled epidemic of final size outcome and found the precision of estimates for threshold parameter improves as the sampling fraction decreases [11]. However, our information is based on temporal data rather than final size outcome it is not sure whether the similar finding will be obtained as in the O'Neill work. The finding on the inverse relationship between the precision of estimates for threshold parameter and the sampling fraction from the O'Neill study [11] would spur us to investigate this issue in the future using samples of temporal individual and household data together with information possibly available from different hierarchical population structures (e.g. schools closures).

In conclusion, the proposed Bayesian hierarchical models are flexible in accommodating the mechanisms of disease transmission, including the Greenwood and the Reed–Frost models, along with factors on an individual level. The models are also capable of taking dependent structures inherited from data on influenza epidemics into account to address the problem of overdispersion when constructing probability models. Although Becker's linear logistic model is capable of testing the Greenwood and Reed–Frost models, it cannot fully account for the characteristics of infectious disease data mentioned above. Hierarchical models with Bayesian underpinnings are one solution to such correlated data with multilevel structures. Applying the proposed Bayesian hierarchical models to the empirical data used to monitor the influenza epidemic demonstrates that the parameters

on the individual, generation, and household levels can be derived based on inferences made using the MCMC method in conjunction with DAG. The Bayesian hierarchical models are also capable of uncovering the interactions between contextual effects, such as socioeconomic status and vaccination, to elucidate their joint effect on the occurrence of influenza. The Bayesian approach is beneficial to the incorporation of prior knowledge on the effect of relevant factors and to the derivation of predictive distributions to assess the evolution of epidemic processes.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mbs.2014.11.006.

References

- [1] N.G. Becker, *Analysis of Infectious Disease Data*, Chapman & Hall/CRC, London, 1989.
- [2] I.M. Longini, J.S. Koopman, Household and community transmission parameters from final distributions of infections in households, *Biometrics* 38 (1) (1982) 115–126.
- [3] F. Ball, D. Mollison, G. Scalia-Tomba, Epidemics with two levels of mixing, *Ann. Appl. Probab.* 7 (1) (1997) 46–89.
- [4] D.J. Daley, J. Gani, *Epidemic Modelling: An Introduction*, Cambridge University Press, Cambridge, 2001.
- [5] T. House, M.J. Keeling, Household, Structure and infectious disease transmission, *Epidemiol. Infect.* 137 (5) (2009) 654–661.
- [6] T. House, M.J. Keeling, Deterministic epidemic models with explicit household structure, *Math. Biosci.* 213 (1) (2008) 29–39.
- [7] F. Ball, O.D. Lyne, Stochastic multi-type SIR epidemics among a population partitioned into households, *Ann. Appl. Probab.* 33 (1) (2001) 99–123.
- [8] F. Ball, P. Neal, A general model for stochastic SIR epidemics with two levels of mixing, *Math. Biosci.* 180 (1–2) (2002) 73–102.
- [9] N. Demiris, P.D. O'Neill, Bayesian inference for epidemics with two levels of mixing, *Scand. J. Stat.* 32 (2) (2005) 265–280.
- [10] N. Demiris, P.D. O'Neill, Bayesian inference for stochastic multitype epidemics in structured populations via random graphs, *J. R. Stat. Soc. Ser. B (Stat. Method.)* 67 (5) (2005) 731–745.
- [11] P.D. O'Neill, Bayesian inference for stochastic multitype epidemics in structured populations using sample data, *Biostatistics* 10 (4) (2009) 779–791.
- [12] S. Cauchemez, F. Carrat, C. Viboud, A.J. Valleron, P.Y. Boëlle, A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data, *Stat. Med.* 23 (22) (2004) 3469–3487.
- [13] H. Brown, R. Prescott, *Applied Mixed Models in Medicine*, second ed., Wiley, Chichester, 2006.
- [14] P.D. O'Neill, D.J. Balding, N.G. Becker, M. Eerola, D. Mollison, Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods, *J. R. Stat. Soc. Ser. C (Appl. Stat.)* 49 (4) (2000) 517–542.
- [15] M. Greenwood, On the statistical measure of infectiousness, *J. Hyg.* 31 (3) (1931) 336.
- [16] W.H. Frost, Some conceptions of epidemics in general, *Am. J. Epidemiol.* 103 (2) (1976) 141–151.
- [17] D.T. Eurich, T.J. Marrie, J. Johnstone, S.R. Majumdar, Mortality reduction with influenza vaccine in patients with pneumonia outside “flu” season: pleiotropic benefits or residual confounding? *Am. J. Respir. Crit. Care Med.* 178 (5) (2008) 527–533.
- [18] E.M. Galarce, S. Minsky, K. Viswanath, Socioeconomic status, demographics, beliefs and A(H1N1) vaccine uptake in the United States, *Vaccine* 29 (32) (2011) 5284–5289.
- [19] T.Y. Ho, K.Y. Huang, T.T. Huang, Y.S. Huang, H.C. Ho, P. Chou, C.H. Lin, C.K. Wei, W.C. Lian, T.C. Chen, H.B. Huang, C.C. Lee, The impact of influenza vaccinations on the adverse effects and hospitalization rate in the elderly: a national based study in an Asian country, *PLoS One* 7 (11) (2012) e50337.
- [20] A.A. Stone, A.B. Krueger, A. Steptoe, J.K. Harter, The socioeconomic gradient in daily colds and influenza, headaches, and pain, *Arch. Intern. Med.* 170 (6) (2010) 570.
- [21] S.R. Shih, G.W. Chen, C.C. Yang, W.Z. Yang, D.P. Liu, J.H. Lin, S.C. Chiu, H.Y. Chen, K.C. Tsao, C.G. Hung, Y.L. Huang, C.K. Mok, C.J. Chen, T.Y. Lin, J.R. Wang, C.L. Kao, K.H. Lin, L.K. Chen, H.L. Eng, Y.C. Liu, P.Y. Chen, J.S. Lin, J.H. Wang, C.W. Lin, Y.J. Chen, J.J. Lu, C.A. Hsiung, P.J. Chen, I.J. Su, Laboratory-based surveillance and molecular epidemiology of influenza virus in Taiwan, *J. Clin. Microbiol.* 43 (4) (2005) 1651–1661.
- [22] F. Carrat, E. Vergu, N.M. Ferguson, M. Lemaître, S. Cauchemez, S. Leach, A.J. Valleron, Time lines of infection and disease in human influenza: a review of volunteer challenge studies, *Am. J. Epidemiol.* 167 (7) (2008) 775–785.
- [23] J. Lessler, N.G. Reich, D.A.T. Cummings, Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school, *N. Engl. J. Med.* 361 (27) (2009) 2628–2636.
- [24] S. Cauchemez, C.A. Donnelly, C. Reed, A.C. Ghani, C. Fraser, C.K. Kent, L. Finelli, N.M. Ferguson, Household transmission of 2009 pandemic influenza A(H1N1) virus in the United States, *N. Engl. J. Med.* 361 (27) (2009) 2619–2627.
- [25] N. Li, G. Qian, R. Huggins, A random effects model for diseases with heterogeneous rates of infection, *J. Stat. Plan. Inference* 116 (1) (2003) 317–332.
- [26] C. Fraser, C.A. Donnelly, S. Cauchemez, W.P. Hanage, M.D. Van Kerkhove, T.D. Hollingsworth, J. Griffin, R.F. Baggaley, H.E. Jenkins, E.J. Lyons, T. Jombart, W.R. Hinsley, N.C. Grassly, F. Balloux, A.C. Ghani, N.M. Ferguson, Pandemic potential of a strain of influenza A(H1N1): early findings, *Science* 324 (5934) (2009) 1557–1561.
- [27] L.M. Sullivan, K.A. Dukes, E. Losina, Tutorial in biostatistics: an introduction to hierarchical linear modelling, *Stat. Med.* 18 (1999) 855–888.
- [28] H. Goldstein, W. Browne, J. Rasbash, Multilevel modelling of medical data, *Stat. Med.* 21 (21) (2002) 3291–3315.
- [29] M.J. Campbell, A. Donner, N. Klar, Developments in cluster randomized trials and statistics in medicine, *Stat. Med.* 26 (1) (2007) 2–19.
- [30] N.G. Best, D.J. Spiegelhalter, A. Thomas, C.E.G. Brayne, Bayesian analysis of realistically complex models, *J. R. Stat. Soc. Ser. A (Stat. Soc.)* 159 (2) (1996) 323–342.
- [31] S.L. Pan, H.M. Wu, A.M.F. Yen, T.H.H. Chen, A Markov regression random-effects model for remission of functional disability in patients following a first stroke: a Bayesian approach, *Stat. Med.* 26 (29) (2007) 5335–5353.
- [32] A.M.F. Yen, H.H. Liou, H.L. Lin, T.H.H. Chen, Bayesian random-effect model for predicting outcome fraught with heterogeneity, *Methods. Inf. Med.* 45 (6) (2006) 631–637.
- [33] D.J. Spiegelhalter, A. Thomas, N.G. Best, Computation on Bayesian graphical models, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), *Bayesian Statistics*, vol. 5, Oxford University Press 1996, pp. 407–425.
- [34] T. Jefferson, C. Di Pietrantonj, L.A. Al-Ansary, E. Ferroni, S. Thorning, R.E. Thomas, Vaccines for preventing influenza in the elderly, *Cochrane Lib.* 3 (2) (2010) 2–33.
- [35] T. Jefferson, C. Di Pietrantonj, A. Rivetti, G.A. Bawazeer, L.A. Al-Ansary, E. Ferroni, Vaccines for preventing influenza in healthy adults, *Cochrane Lib.* 4 (7) (2010) 1–20.
- [36] D.J. Lunn, A. Thomas, N. Best, D.J. Spiegelhalter, Winbugs—a Bayesian modelling framework: concepts, structure, and extensibility, *Stat. Comput.* 10 (4) (2000) 325–337.
- [37] T. House, N. Inglis, J. Ross, F. Wilson, S. Suleman, O. Edeghere, G. Smith, B. Olowokure, M. Keeling, Estimation of outbreak severity and transmissibility: influenza A(H1N1)pdm09 in households, *BMC Med* 10 (2012) 115–117.
- [38] R.E.H. Simpson, I. Sutherland, Does influenza spread within the household? *Lancet* 266 (6814) (1954) 721–726.
- [39] S. Cauchemez, A. Bhattarai, T.L. Marchbanks, R.P. Fagan, S. Ostroff, N.M. Ferguson, D. Swerdlow, t.P.H.N.w. group, Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza, *Proc. Natl. Acad. Sci. U.S.A.* 108 (7) (2011) 2825–2830.
- [40] D. Zhang, W. Liu, P. Yang, Y. Zhang, X. Li, K.E. Germ, S. Tang, W. Sun, Q. Wang, Factors associated with household transmission of pandemic (H1N1) 2009 among self-quarantined patients in Beijing, China, *PLoS One* 8 (10) (2013) e77873.
- [41] J.C. Thomas, D.J. Weber, *Epidemiologic Methods for the Study of Infectious Disease*, Oxford University Press, New York, 2001.
- [42] Q. Gu, N. Sood, Do people taking flu vaccines need them the most? *PLoS One* 6 (12) (2011) e26347.
- [43] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, *Bayesian Data Analysis*, second ed., Chapman & Hall/CRC, London, 2004.
- [44] P. Armitage, G. Berry, J.N.S. Matthews, *Statistical Methods in Medical Research*, fourth ed., Blackwell Science, Massachusetts, 2002.
- [45] L.Y. Chang, W.H. Chen, C.Y. Lu, P.L. Shao, T.Y. Fan, A.L. Cheng, L.M. Huang, Household transmission of pandemic (H1N1) 2009 virus, Taiwan, *Emerg. Infect. Dis.* 17 (10) (2011) 1928–1931.
- [46] A. Glatman-Freedman, I. Portelli, S.K. Jacobs, J.J. Mathew, J.E. Slutzman, L.R. Goldfrank, S.W. Smith, Attack rates assessment of the 2009 pandemic H1N1 influenza in children and their contacts: a systematic review and meta-analysis, *PLoS One* 7 (11) (2012) e50228.
- [47] R.E. Thomas, Is influenza-like illness a useful concept and an appropriate test of influenza vaccine effectiveness? *Vaccine* 32 (19) (2014) 2143–2149.
- [48] J.W. Jian, G.W. Chen, C.T. Lai, L.C. Hsu, P.J. Chen, S.H.S. Kuo, H.S. Wu, S.R. Shih, Genetic and epidemiological analysis of influenza virus epidemics in Taiwan during 2003 to 2006, *J. Clin. Microbiol.* 46 (4) (2008) 1426–1434.
- [49] J.H. Chuang, A.S. Huang, W.T. Huang, M.T. Liu, J.H. Chou, F.Y. Chang, W.T. Chiu, Nationwide surveillance of influenza during the pandemic (2009–10) and post-pandemic (2010–11) periods in Taiwan, *PLoS One* 7 (4) (2012) e36120.
- [50] Y.C. Lo, J.H. Chuang, H.W. Kuo, W.T. Huang, Y.F. Hsu, M.T. Liu, C.H. Chen, H.H. Huang, C.H. Chang, J.H. Chou, Surveillance and vaccine effectiveness of an influenza epidemic predominated by vaccine-mismatched influenza b/yamagata-lineage viruses in taiwan, 2011–12 season, *PLoS One* 8 (3) (2013) e58222.
- [51] A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy, J. Schweinle, Clinical signs and symptoms predicting influenza infection, *Arch. Intern. Med.* 160 (21) (2000) 3243–3247.
- [52] K. Thursky, S.P. Cordova, D. Smith, H. Kelly, Working towards a simple case definition for influenza surveillance, *J. Clin. Virol.* 27 (2) (2003) 170–179.