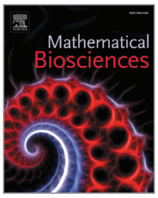
Mathematical Biosciences 261 (2015) 13–26

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs



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.

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).

http://dx.doi.org/10.1016/j.mbs.2014.11.006
0025-5564/

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, $\geq$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 $\leq$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