Epidemics 7 (2014) 7–12

Contents lists available at ScienceDirect

Epidemics

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

An approach for modeling cross-immunity of two strains, with application to variants of *Bartonella* in terms of genetic similarity

Kwang Woo Ahn[a,*], Michael Kosoy[b], Kung-Sik Chan[c]

[a] Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, USA
[b] Centers for Disease Control and Prevention, Fort Collins, CO, USA
[c] Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA, USA

ARTICLE INFO

Article history:
Received 21 February 2012
Received in revised form 14 March 2014
Accepted 14 March 2014
Available online 24 March 2014

Keywords:
Bartonella
Conditional least squares
Cross-immunity
SIR model

ABSTRACT

We developed a two-strain susceptible-infected-recovered (SIR) model that provides a framework for inferring the cross-immunity between two strains of a bacterial species in the host population with discretely sampled co-infection time-series data. Moreover, the model accounts for seasonality in host reproduction. We illustrate an approach using a dataset describing co-infections by several strains of bacteria circulating within a population of cotton rats (*Sigmodon hispidus*). *Bartonella* strains were clustered into three genetically close groups, between which the divergence is correspondent to the accepted level of separate bacterial species. The proposed approach revealed no cross-immunity between genetic clusters while limited cross-immunity might exist between subgroups within the clusters.

Introduction

Multi-strain models have been widely used in epidemiology (Gupta et al., 1998; Kamo and Sasaki, 2002; Abu-Raddad and Ferguson, 2005; Bianco et al., 2009; Minayev and Ferguson, 2009). Developing and using multi-strain models is a challenging procedure due to numerous parameters such as death rate, birth rate, force of infection, and transmission rate, which are commonly assumed to be strain specific.

One of the key concepts of these models is cross-immunity, which allows infection by one strain to induce partial/perfect protection against other strains. Gupta et al. (1998) proposed a very general model accounting for the cross-immunity in a multi-strain system, based on which they studied the effects of cross-immunity on evolution of strain structure. Abu-Raddad and Ferguson (2005) investigated population dynamics of host-pathogen systems involving an arbitrary number of antigenically distinct strains whose interaction depends on the cross-immunity. Minayev and Ferguson (2009) studied multi-strain deterministic epidemic models in which cross-immunity varies with the genetic distance between strains. Kamo and Sasaki (2002) proposed a two-strain susceptible-infected-recovered (SIR) model with cross-immunity. These models, however, assume an equilibrium population size over time, i.e., equal, constant birth and death rates. These assumptions might be too strong since the host population might fluctuate dramatically between seasons, which may affect the force of infection (Davis et al., 2005). For example, hispid cotton rat populations usually have peak litter production occurring in late spring and in late summer-early fall (Cameron and Spencer, 1984).

In addition, these earlier works were restricted to simulation studies assuming known parameter values. Another issue is that, except for the model of Gupta et al. (1998), the state variables of these SIR-based models are often expressed in terms of the densities of various categories of the hosts. In general, it is often difficult to estimate the number of susceptibles, infectives, and recovered subjects over time, which makes an application of such models to real data challenging. Instead, developing a model consisting of proportions of susceptibles, infectives, and recovered subjects may make data analysis more feasible. Furthermore, some of state variables may not be observable in practice, for example, only infected individuals may be identified. Thus, it is pertinent to develop appropriate statistical models with partially observed data.

In this paper, we propose a two-strain SIR model that extends the model of Gupta et al. (1998) and that of Kamo and Sasaki (2002) by accounting for seasonality in host reproduction and non-constant death rate. Furthermore, the proposed model is applied to a real dataset monitoring on the prevalence of co-infections of several *Bartonella* strains in a natural population of cotton rat

* Corresponding author at: Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI 53226, USA. Tel.: +1 414 955 7387; fax: +1 955 6513.
E-mail addresses: kwooahn@mcw.edu (K.W. Ahn), mck3@cdc.gov (M. Kosoy), kung-sik-chan@uiowa.edu (K.-S. Chan).

http://dx.doi.org/10.1016/j.epidem.2014.03.001

(*Sigmodon hispidus*). This dataset is referred to as the Bartonella data and is described in the next section. Details of the proposed two-strain SIR model and the statistical methods are elaborated in the Two-strain SIR model with state variables expressed as proportions section and Method section, respectively. Interpretation and discussion of the epidemiological significance of the analysis results are given in the Results section. A brief conclusion is given in the last section.

Data description

The field data used for this analysis were collected from a longitudinal study that monitored the prevalence of *Bartonella* infection in a wild cotton rat population near Social Circle, Walton County, GA, USA, over a period of 17 months, from March, 1996 to July, 1997, except December 1996, yielding altogether 483 trapping records (Kosoy et al., 2004a,b). Cotton rats were captured for two or three consecutive nights each month and blood samples were taken. First-time captured cotton rats were marked. Marked and sampled rats were released. Sixty four out of 483 trapped rats were found to have co-infections by two or three *Bartonella* strains. Based on the cluster analysis of the genetic sequences among the bacterial isolates obtained from cotton rats (*S. hispidus*) in Georgia, identified *Bartonella* strains were clustered into three genogroups based on the similarities of the *gltA* sequences: A, B, and C (similarity range 88.2–93.5%). The citrate synthase gene, *gltA*, is a popular and widely used target to distinguish between closely related *Bartonella* species and genotypes (Kosoy et al., 2012). Since (Norman et al., 1995) proposed the use of a variable fragment of this gene to differentiate *Bartonella*-like isolates at the species level, most laboratories working with *Bartonella* bacteria have successfully applied this genetic marker. (Birtles and Raoult, 1996) have also demonstrated that the gltA-derived phylogeny appears to be more useful than the phylogeny derived from 16Sr DNA sequence data for investigating the evolutionary relationships of *Bartonella* species.

The three genogroups were further classified into unique sequence strains A1–A5, B1–B5, and C1–C2 with the sequence similarity ranged from 96.2% to 99.7%; see Table 1 of Kosoy et al. (2004b). In June and July of 1996, four cotton rats gave birth in their traps. To avoid issues related to vertical transmission of *Bartonella* infection from parent subjects to their children, we excluded 19 neonatal rats captured in June and July. Fig. 1 shows the time-series plots of (i) the monthly numbers of trapped cotton rats (bottom figure) and (ii) proportions of trapped cotton rats that were infected by each *Bartonella* strain (top figure), which shows that A1 was the dominant strain and the prevalence of strain B was low. In this paper, we consider two scenarios of cross-immunity: (i) between genogroups; (ii) between variants in the same genogroup. For the first scenario, we combined B and C due to low frequency of strain B and relatively high genetic similarity between strains B and C (Table 1 of Kosoy et al., 2004b). For the second scenario, we consider genogroup A only because of its high prevalence. Table 1 of Kosoy et al. (2004b) shows that A1 and A5 are genetically close to each other, and so are A2 and A4. Therefore, in this report, we compare A1&A5 vs. A2&A4 and A vs. B&C.

A two-strain SIR model with state variables expressed as proportions

We consider the two-strain special case of the multi-strain model proposed by Gupta et al. (1998). Their model provides a general framework for modeling the dynamics of an infectious disease with multiple strains of a pathogen that may induce various degrees of cross-immunity in the hosts. Here, we extend their model to allow for variable host reproduction, and that the death rate can also be variable and not equal to the birth rate. Moreover, we modify the model so that a host is assumed to only make a fixed number of contacts with other hosts, on average. Detailed derivation of the model is given in the Supplementary Material.

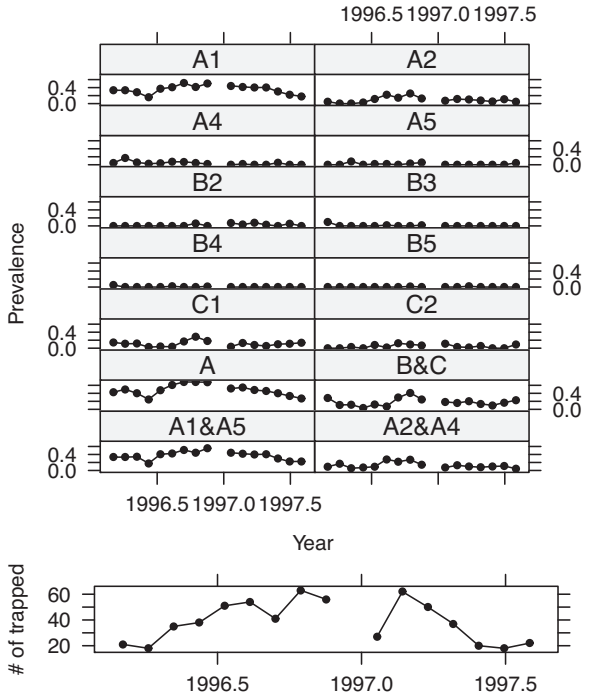

Fig. 1. Prevalence of *Bartonella* strains and the number of trapped cotton rats over the study period. Solid circles show observed values. The *y*-axis of the bottom figure represents the number of trapped cotton rats.

The five state variables are: $x = x_{SS}$, $y_1 = x_{I\bullet}$, $y_2 = x_{\bullet I}$, $z_1 = x_{IS} + x_{RS}$, $z_2 = x_{SI} + x_{SR}$, where all variables are proportions of hosts with particular disease status indicated by the double subscripts with the first subscript being *S*, *I*, *R*, standing for susceptible to the first strain, infected by the first strain, recovered from an infection by the first strain, and a subscript is set to $\bullet$ if no condition is imposed on the particular strain; the second subscript refers to the disease status with respect to the second strain. For example, x_{IS} is the proportion of hosts infected by the first strain but susceptible to the second strain, $x_{I\bullet}$ is the proportion of hosts infected by the first strain, while $x_{\bullet I}$ is the proportion of hosts infected by the second strain. All state variables are implicit functions of time t with their derivatives denoted by the dot notation. The extended two-strain SIR model is given as follows:

$$\begin{aligned}
\dot{x} &= -\alpha_1 x y_1 - \alpha_2 x y_2 + (1-x)b,\\
\dot{y}_1 &= \alpha_1(x + \delta z_2)y_1 - (\gamma_1 + b)y_1,\\
\dot{y}_2 &= \alpha_2(x + \delta z_1)y_2 - (\gamma_2 + b)y_2,\\
\dot{z}_1 &= \alpha_1 x y_1 - \alpha_2 \delta z_1 y_2 - b z_1,\\
\dot{z}_2 &= \alpha_2 x y_2 - \alpha_1 \delta z_2 y_1 - b z_2,
\end{aligned} \tag{1}$$

where the parameter α_i's are the transmission rates between an individual infected by strain i ($i = 1, 2$) and one susceptible to both strains, γ_i's are the host's recovery rate from an infection by strain i, δ is the cross-immunity parameter, and $b = b_t$ is the birth rate. Note that the death rate $\mu = \mu_t$ is eliminated in the algebra so that it no longer appears in (1). In other words, the death rate does not affect the dynamics when the state variables are expressed as proportions, i.e., $(x, y_1, y_2, z_1, z_2)^T$ in this model. Thus, (1) makes it feasible to analyze the general two-strain system without the need to know or to estimate μ. The non-negative parameter δ controls