Modelling Immunity Against Infectious Plague Bacteria in Great Gerbils in Kazakhstan: a Continuous-time Markov Chain Approach

Siyun Park,[1] Kung-Sik Chan,[1,*] Hildegunn Viljugrein,[2] Larissa Nekrassova,[3] Bakhtiyar Suleimenov,[3] Vladimir S. Ageyev,[3] Nikolay L. Klassovskiy,[3] Sergey B. Pole[3] and Nils Chr. Stenseth[2]

[1] Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, U.S.A.

[2] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, P.O. Box 1050 Blindern, N-0316 Oslo, Norway

[3]M. Aikimbaev's Kazakh Scientic Center for Quarantine and Zoonotic Diseases, 14 Kapalskaya Street, Almaty 480074, Republic of Kazakhstan.

September 1, 2004

Summary. We propose a new stochastic framework for analyzing the dynamics of the (acquired) immunity of wildlife hosts against organisms causing an infectious disease. Our study is motivated by the need for analyzing monitoring time-series data covering the period from 1975 to 1995 on bacteriological and serological tests-samples from great gerbils being the main

[*] *email:* kchan@stat.uiowa.edu

host of *Yersinia pestis* in Kazakhstan. Based on a four-state continuous-time Markov chain, we derive a generalized nonlinear mixed-effect model for analyzing the serological test data. The immune function of a host pertains to its production and activation of cells that fight infection. We find that the immune function of the sampled great gerbils is seasonal so as to make the great gerbils lose immunity to plague faster over the winter-to-summer season than over the summer-to-winter season.

Key words:
binomial distribution, continuous-time Markov chain, generalized nonlinear mixed-effect model, panel of time series

1. Introduction

We consider the problem of studying the dynamics of the immune responses of wildlife hosts to infectious disease, based on monitoring time-series data. In monitoring epizootics of an infectious disease among wildlife hosts, the sampling scheme is often destructive as the sampled hosts may have to be sacrificed in order to test for the presence of the agents causing the infectious disease, and including testing for the presence of anti-bodies in blood and/or inner organs of the hosts. Such studies, therefore, cannot provide information on the time course at the population level of the immune response of an individual host. Our motivating case study concerns the modelling of the dynamics of the immune responses of great gerbils (*Rhombomys opimus*) to plague in Kazakhstan. The great gerbil populations constitute several natural foci to plague (caused by the bacteria *Yersinia pestis*) in Kazakhstan where the disease may be transmitted to humans by vectors, mainly, fleas. A long-term monitoring study of this natural plague system was undertaken

from 1949-1995, for tracking the prevalence of plague in the great gerbil population; see Davis *et al.* (2004) and Frigessi *et al.* (2004). In particular, in both spring and fall of each year, samples of great gerbils provided bacteriological and serological test data for plague symptoms. While bacteriological tests may detect the presence of plague bacteria and hence the plague disease in great gerbils at the time of sampling, serological tests may detect the presence of antibodies to plague bacteria. Consequently the serological test data are indicative of past infections, and may shed light on the dynamical structure of the immune system of the great gerbils against plague; see Section 3.

In this paper, we focus on the modeling of the dynamics of the (acquired) immunity of great gerbils in Kazakhstan against plague. For this purpose, we assume a continuous-time Markov chain model (Cox and Miller, 1968, and Bhattacharya and Waymire, 1990) for the time course of the immune response of a (random) great gerbil with a 4-state state space; the four states are susceptible (S), infected (I), recovered (R) and death (D). An advantage of this approach is its relative ease of deriving an approximate expression relating the probability of a positive serological test with the probability of a positive bacteriological test and its lags. Consequently, we can use this relationship to study the dynamics of the immune response (to plague) of the great gerbils, conditional on the estimated prevalence rates of plague disease among the great gerbil population. A drawback of our continuous-time Markov chain approach is that (unlike the classical approaches to infectious disease modeling; see Grenfell and Dobson, 1995, and Dickmann and Heesterbeek, 2000 for recent surveys) it does not model the interactions between the

infected hosts, the recovered hosts and the susceptible hosts. This omission is, however, deliberate and it simplifies the analysis; this is appropriate as our goal is not to model the epidemiological time course, but to understand the dynamics of the immune response, conditional on the prevalence rates. We specify the instantaneous prevalence rate to be constant over each season, but otherwise be a free parameter. Based on the continuous-time Markov chain framework, we derive a generalized nonlinear mixed-effect model for analyzing the serological test data. For recent surveys of generalized nonlinear mixed-effect models, see, for instance, Pinheiro and Bates (1995) and Davidian and Giltinan (2003).

This article is organized as follows. In Section 2 we provide some general background information about the plague system as it is found in Kazakhstan. In Section 3 we provide further details on the monitoring data, the protocol of the bacteriological and the serological tests. In Section 4 we derive a generalized nonlinear mixed-effect model for analyzing the serological test data. Details of the derivation are collected in an appendix. Section 5 presents the results of the data analysis. We conclude in Section 6.

2. The plague system

Plague exists in nature as a disease of wild rodents caused by infection of the bacterium, *Yersinia pestis*. The infection is maintained in natural foci of the disease in wild rodent colonies through transmission between rodents by their flea ectoparasites. Rodents are the primary hosts of *Y. pestis*; however, other mammals, including humans, may be infected. Other mammalian species, especially rodent-consuming carnivores, may play an ecologically important role by transporting infected fleas from one area to another. Plague

is widespread on all continents of the world except Australia and Antarctica. In the desert and semidesert of Kazakhstan (and central Asia, in general), *Rhombomys opimus* (the great gerbil) and the fleas inhabiting their burrows (mainly of the genus *Xenopsylla*) are considered to be the main host and vectors, respectively, of plague.

Great gerbils are social desert rodents living in family groups that occupy discrete, permanent burrow systems (colonies). A family group usually includes a single adult male and several females with their offspring. Group size and composition vary by year with variation in population densities. Females remain in their natal group, while males disperse and join other solitary females or female groups. Both males and females defend the group; males usually chase away other males. The winter death rate is high; no more than 10-12 percent of adult gerbils survive the winter. All members of a family group participate in the storing of green vegetation. Food storage is especially intensive in the spring (when pups emerge) and in fall. From one to three litters, ranging in size from 1-14 (but usually 4-7), are born from April to September, depending on precipitation and vegetation (Naumov and Lobachev, 1975). Great gerbils exhibit extensive population fluctuations between years. Years of high abundance seems to be necessary for the large scale spread of plague that produces epizootics (Davis *et al.*, 2004).

The response of great gerbils to plague infection depends on the intensity of the infection (i.e., quantity and virulence of plague microbes) as well as the level of innate and acquired resistance against the plague infection. The higher the level of resistance, the weaker the infectious process may be (Rothschild, 1978). Great gerbils are relatively resistant to plague, but there

is extensive individual variation. Resistance to plague may differ dramatically among populations of the same species of rodent, and the rate may change within populations depending on how recently they were exposed to plague. Resistant great gerbils rarely die of plague, but do become bacteremic (i.e., with bacteria present in the blood) and, therefore, may serve as sources of infectious blood meals for feeding fleas. The rodents are infectious only in the period of bacteremia of any intensity. The incubation period is 3-5 days. Great gerbils with an acute plague infection are characterized by reduced activity being more vulnerable to predation. Infected great gerbils experience higher death rates than non-infected great gerbils.

3. The Monitoring Data from Kazakhstan

The survey area is located south-east of Lake Balkhash in south-eastern Kazakhstan, being part of the PreBalkhash plague focus (see, e.g., Davis *et al.*, 2004). The PreBalkhash focus is separated into specific Landscape-Epizootological Regions, a Landscape-Epizootological Region being an area of a particular type of landscape, soil, vegetation, density of the gerbils and their fleas, as well as by the level of epizooticity. For the purposes of monitoring plague, the whole of Kazakhstan was divided into 40×40 km squares, here referred to as large squares (see Figure 1). Each large square compromises 4 20×20 km primary squares which in turn are divided into 4 sectors. Within a given sector, data are typically recorded twice a year and providing information on the results of bacteriological and serological tests (prevalence data) together with independent information on the rodent densities.

The sampling was done during spring and fall from 1949 to 1995. The great gerbils are mainly caught around May-June in spring and around

September-October in fall, with the bacteriological and serological tests administered to the caught gerbils. Gerbil density data are estimated in April and September. For simplicity, it is assumed that the sampling was carried out at time Δ_t in the tth season which spans from b_t to L_t. Standardized serological tests have been carried out since 1975; hence, the analysis reported below are confined to the data collected from the period starting from 1975. The bacteriological test is positive when *Y. pestis* is isolated by planting rodent samples (blood, liver and spleen) on agar media. The result of the bacteriological test is the result of the combined testing of blood, liver and spleen samples taken from each single rodent. Bacteriologically positive results are found for rodents with local forms of infection in organs (liver or spleen) or with expressed bacteremia (i.e., live bacteria present in blood).

The more *Y. pestis* there are in the sample, the more likely there will be a positive test result. However, samples taken from infected individuals may not include live bacteria, because the likelihood for bacteria to be included in the sample is dependent on the strength of the infection (the amount and the virulence of bacteria), the great gerbil's resistance to plague and the phase of the infectious process in the great gerbil caught. Hence, not all infected gerbils will give a positive response to the bacteriological test. Additionally, the reduced above-ground activity of plague-infected great gerbils makes them harder to trap than non-infected individuals. This implies that the probability that a caught great gerbil is bacteriologically positive approximately equals the probability that a (random) great gerbil is bacteriologically positive up to a multiplicative factor that is between 0 and 1; the multiplicative factor quantifies the reduction in the trapping probability due to the lower

mobility and higher mortality rate of an infected great gerbil. The multiplicative factor is assumed to be constant or seasonal. Consequently, the fraction of bacteriologically positive caught great gerbils is a biased estimator of the prevalence rate, with a downward bias. Fortunately, the (unknown) multiplicative factor can be absorbed by a parameter in the statistical model to be derived below, so that, without loss of generality, the multiplicative factor can and will be taken to be 1; see Section 4. In other words, it is valid to carry out the analysis conditional on the (biased) prevalence estimates, as it is done in Section 5.

We remark that the Markov chain model used in this paper only attempts to model the average infectivity, immunity and survival rates as the parameters in the transition rate matrix are assumed to be unknown, but fixed, parameters. As discussed above, some of these epidemiological parameters may be better modelled as random; however, the data currently under study is too coarse for such a purpose.

Serological tests were done by analyzing blood samples and tissue extracts from the caught great gerbils for the presence of anti-F1 antibodies by a passive haemagglutination (PHA) test. The F1 antigen is specific to *Y. pestis* (Perry and Fetherston, 1997). The specificity of the PHA test was confirmed by an F1 antigen neutralization test, reducing the probability of a false positive test result.

When infected great gerbils develop antibodies to *Y. pestis*, they "seroconvert" from antibody-negative to antibody-positive. After an infectious flea bite, antibodies with low titres (i.e., low concentrations) may start to be registered from the third day. Complete antibodies (both IgG and IgM gam-

maglobulins), are commonly registered after 6-8 days in the rodent (Kanatov *et al.*, 1969). The probability of a positive serological result depends on the length of time between the rodent was infected with the bacteria and was caught; the shorter the time period since seroconversion in the rodent, the higher the likelihood of a positive test result. After seroconversion, production of antibodies may increase for a few weeks, thereafter the antibody titres usually diminish gradually over the next months. The average length of antibody registration is 3-6 months, but in some cases it may last up to one year (Kanatov, 1974) and possibly longer. From autumn to spring, antibodies are found in a minority of great gerbils (10%; Suleimenov *et al.*, 2001).

Positive serological tests indicate recent or past exposure to plague. Negative serological tests are not evidence for the absence of plague, because individual rodents may not react serologically or have very low diagnostic titres (corresponding to a low infectious dose, initial or final stages of antibody production and/or low reactivity of an individual). Intensity of antibody production (and the length of their presence) depends on immune reactivity of a rodent. The later the elimination of the plague bacteria, the clearer is the immune response of the rodent (Kanatov, 1974).

The data used in this paper consist of counts of great gerbils, the numbers of great gerbils that are bacteriologically positive in each sample, as well as the counts of serologically positive great gerbils. We analyze data from 6 large squares that have adequate data for the model estimation reported in Section 5. The time-series plots of counts of serologically positives, and those of the bacteriologically positives, the sample sizes and the rodent density estimates are depicted in Figure 1. These plots display a variety of temporal patterns,

and diverse levels of serological and bacteriological positive counts. Since 1975, most large squares had at most two epizootics.

For the purpose of analyzing these data, we formulate a new model framework for analyzing such data. We aim to explain the observed variation in the dynamical patterns in terms of an epidemiological-based model, and to improve our understanding of the underlying immunity loss structure of the great gerbil system in Kazakhstan.

4. A Generalized Nonlinear Mixed-effect Model

Given the sample size, the number of bacteriological positives and that of serological positives have approximately marginal binomial distributions. Here, we focus on the analysis of the serological test data. For this purpose, we first derive a formula for the probability, q_t, that a great gerbil is serologically positive at time Δ_t, the tth sampling epoch. (Recall that the tth sampling season begins at b_t and ends at L_t with the sampling epoch Δ_t within the season.) Specifically,

$$q_t = C + \sum_{j=0}^{M} \frac{D_{t-j}}{D_t} \tau_{t-j} p_{t-j} \prod_{k=1}^{j} \theta_{t-k+1} + \eta_t, \tag{1}$$

where a product over an empty index set is defined to be 1; D_t is the estimated great gerbil density at time Δ_t; τ_t is proportional to the instantaneous recovery rate (the proportionality constant is, however, generally region-specific); p_t is proportional to the probability of a great gerbil being bacteriologically positive at time Δ_t, i.e., the instantaneous plague prevalence rate at time Δ_t; the maximum value of M is set to be 7 because a great gerbil rarely lives longer than 4 years (Naumov and Lobachev, 1975); θ_t is the probability of a recovered great gerbil keeping immunity throughout the tth season (i.e.,

from b_t to L_t); C is a term accommodating for possible migration of great gerbils between neighboring squares, and η_t are independent $N(0, \sigma^2)$ errors. The plague prevalence rates are subject to no restrictions and can vary from season to season, and year to year.

Consider the probability of a great gerbil being bacteriologically positive at time Δ_t, which can be obtained as follows (see the Appendix for details).

$$\begin{aligned} p_t &= P(\text{a great gerbil is bacteriologically positive at } \Delta_t) \quad (2) \\ &\approx q_{SI} e^{q_{II}(\Delta_t - b_t)} \frac{e^{(q_{SS} - q_{II})(\Delta_t - b_t)} - 1}{(q_{SS} - q_{II})}, \quad (3) \end{aligned}$$

where q_{SI} is the (instantaneous) infective rate, $(-q_{SS})^{-1}$ is the mean susceptible period, and $(-q_{II})^{-1}$ is the mean infective period.

Because a plague-infected great gerbil is harder to be caught, the probability p_t^T that a caught great gerbil is bacteriologically positive at Δ_t is different from p_t. Their relationship can be derived by Bayes' theorem. Let ω be the (conditional) probability of a great gerbil being caught given that it is healthy and $\delta\omega$ be the (conditional) probability of a great gerbil being caught given that it is plague-infected, where for $0 < \delta \leq 1$. By Bayes Theorem,

$$p_t^T = \frac{p_t \delta \omega}{p_t \delta \omega + (1 - p_t)\omega} \approx \delta p_t,$$

where the approximation error is of the order of p_t^2. Similarly, it can be

shown that

$$\begin{aligned} q_t &= P(\text{a great gerbil is serologically positive at } \Delta_t) \\ &\approx \sum_{j=0}^{M} \frac{N_{t-j}}{N_t} \prod_{k=1}^{j} \theta_{t-k+1} \times p_{t-j} \times \tau_{t-j}, \qquad (4) \\ &= \sum_{j=0}^{M} \frac{N_{t-j}}{N_t} \prod_{k=1}^{j} \theta_{t-k+1} \times p_{t-j}^{T} \times (\frac{\tau_{t-j}}{\delta}), \end{aligned}$$

where N_t is the true number of great gerbils in the tth season. But δ can be absorbed into τ_{t-j}. For simplicity, we henceforth assume $\delta = 1$ and $p_t^T = p_t$. Furthermore, it is assumed that τ_t and/or θ_t are seasonal, i.e.,

$$\tau_t = \begin{cases} \tau_S, & \text{if } t \text{ is a spring (winter-to-summer season),} \\ \tau_F, & \text{if } t \text{ is a fall (summer-to-winter season),} \end{cases}$$

with θ_t similarly specified. Later on, we shall consider the further simplification that θ_t and/or τ_t are constants.

Because N_t are unknown, they will be replaced by great gerbil density estimates. Specifically let D_t be the estimated great gerbil density. Then a simple model relating the density estimates to N_t is: $d_t = K + n_t + \epsilon_t$, where K is a constant, $d_t = \log(D_t)$, $n_t = \log(N_t)$ and ϵ_t are iid $N(0, \sigma^2)$ noises. Consequently,

$$\frac{D_{t-j}}{D_t} = \exp(n_{t-j} - n_t) \exp(\epsilon_{t-j} - \epsilon_t) = \frac{N_{t-j}}{N_t} \exp(\epsilon_{t-j} - \epsilon_t).$$

This implies that

$$q_t = \sum_{j=0}^{M} \frac{D_{t-j}}{D_t} \tau_{t-j} p_{t-j} \prod_{k=1}^{j} \theta_{t-k+1} \exp(\epsilon_t - \epsilon_{t-j}).$$

The rather complex noise structure can be simplified by noting that (i) for small noise terms, $\exp(\epsilon_t - \epsilon_{t-j})$ can be approximated by $1 + (\epsilon_t - \epsilon_{t-j})$ and (ii)

then approximating the combined noise terms as a single stochastic error, i.e., we specify (1). (Strictly speaking, the error terms in (1) are heteroscedastic, but are assumed to be homoscedastic for simplicity.)

The probability model introduced above provides a useful framework for analyzing the bacteriological and serological test data. While the bacteriological test data provide only information on the plague prevalence, the serological test data contain information on plague prevalence as well as the immunity loss structure. Thus, ideally, a joint analysis may be carried out with both the bacteriological and serological test data. However, such a joint analysis is limited by the fact that there were just a few epizootics in each large square making a fair number of p_t being zero. For the sake of comparison, the analysis were carried out for data starting from spring 1975, for the 6 squares under study. The lag structure of p_t entering in q_t defined in (1) further complicates the joint analysis. Because of these difficulties we first estimate p_t based on the bacteriological test data. Recall that in the tth season, a sample of G_t great gerbils were given bacteriological and serological tests for plague symptoms. Let B_t denote the number of bacteriological positives and S_t the number of serological positives in the sample. Below, the notation $B(n, p)$ denotes the Binomial distribution with n trials and the probability of success being p. Specifically, because B_t have the Binomial distributions, $B(G_t, p_t)$, the maximum likelihood estimates of p_t equal $\hat{p}_t = B_t/G_t$. (Recall that we have made the simplifying assumption that $p_t = p_t^T$. Hence, more generally, $\hat{p}_t^T = B_t/G_t$). We then treat $\hat{p}_t$ as if they were the true p_t and proceed to analyze the serological test data, for those seasons since spring of 1975 and with non-zero serologically positives. That is, we model S_t given

G_t as $B(G_t, q_t)$, where q_t is given by (1) with p_t there replaced by $\hat{p_t}$. This is a generalized nonlinear mixed-effect model, and the models reported in the next Section were fitted using proc nlmixed of SAS. The estimation was done with the parameters τs parameterized to be positive and Cs and θs be between 0 and 1; furthermore we truncate the right side of (1) if it falls below 0 or above 0.99.

5. Statistical Analysis and Discussion

We considered four cases, namely, case 1: constant τ and seasonal θ, case 2: seasonal τ and constant θ, case 3: both τ and θ are constants and case 4: both τ and θ are seasonal. The analysis are done with the data from large squares 91, 93, 105, 106, 117 and 118 (see Figure 1). The other large squares do not have adequate serological positive data for analysis, and hence are omitted. Given the relatively few epizootics, we assume a common θ and τ model for all 6 large squares, but the constant C is allowed to be square specific. We restrict M to be ≤ 7, which amounts to a maximum of great gerbil life-span of 4 years. Table 1 reports the AIC_C of the fitted models for the four cases, with various values of M in the model defined by (1). The minimum AIC_C is bold-faced, while entries for models whose Hessian matrices are singular or non-negative-definite are printed in italics. Thus, a seasonal θ and constant τ model with $M = 4$ is selected, based on AIC_C, as the best model with common immunity structure. That is, the best model chosen includes a 4-seasonal-lag disease structure where the immunity loss rate differs between the winter-to-summer season and the summer-to-winter season, while the recovery rate is constant between the two seasons. The coefficient estimates and some statistics related to the estimates of the best

model are summarized in Table 2. The estimates of the main parameters, θ_F, θ_S, τ and σ^2 are significant.

The estimate of the parameter θ is significantly higher in fall than in spring, suggesting that recovered great gerbils lost immunity to plague more quickly during the winter-to-summer season than the summer-to-winter season. In PreBalkash area, reproduction occurs mainly over April-June (Naumov and Lobachev, 1975). The fact that the period of intensive reproduction being in the winter-to-summer season is likely to be one of the reasons for the estimate of the parameter θ being lower in spring. Immunity is generally compromised during the breeding compared to the non-breeding season (Nelson *et al.*, 2002). Mounting an immune response likely require resources that could otherwise be allocated to other biological functions. Nelson *et al.*, (2002) predict that immune function should be reduced when energetic requirements are high (e.g., during migration, pregnancy, territory defence, lactation or winter). As winters progress and nutritional and climatic stressors act on the great gerbils, we may expect declines in body mass and levels of immunocompetence during late winter and early spring.

We have checked the assumption of common structure by fitting the unconstrained model to each large square. However, the model fits to large squares 93 and 105 are problematic (e.g., having non-definite Hessian matrices, which is probably due to the low number of positive serological cases). Hence, a fair comparison between the constrained (common-structure) model and the unconstrained model is not possible. Indeed, this is the main reason for imposing the common structure assumption.

The immigration effects are insignificant for large squares 91, 105 and

117, but they are significant for large squares 93, 106 and 118. Indeed, migration of great gerbils from other large squares into the latter three squares may account for the surge in the serological positive cases toward the end of the sampling period for these three squares; see Figure 1. Moreover, while for large squares 91, 105 and 117, the bacteriological time series always led the serological series, this was not the case for large squares 93, 106 and 118, and instead the serological series sometimes led the bacteriological series; immigration may be a likely cause of this phenomenon. (An alternative explanation is that the lower trapping probability of the sick animals (likely to give rise to a positive bacteriological test result) might contribute to the pattern of serological series sometimes leading the bacteriological series; more field work on the relative magnitude of the differential trapping probabilities is needed to investigate this scenario.) Large square 91 belongs to Landscape-Epizootological Region Bakanas, large squares 105 and 117 mainly belongs to Landscape-Epizootological Region Akdala, while large squares 93, 106 and 118 consist of a combination of data from both Akdala and Landscape-Epizootological Region Saryesikotrau. These last three squares are exactly those squares for which an migration effect is found. To assess the possible problem due to pooling data from different Landscape-Epizootological Regions, we computed the data according to Landscape-Epizootological Region and repeated the analysis for each region; we found that while the seasonal immunity structure is still supported by data from Akdala, data from the other two regions select the constant θ and τ model. However, Akdala has 58% of all serological positives that occurred in the three Landscape-Epizootological Regions. The low sample size in the other two regions may

weaken the embedded seasonal signal. Consequently, it seems to us that the combined analysis by large squares is still the best way to summarize the data. Hence, we conclude that it is appropriate to combine data from the three different Landscape-Epizootological Regions to estimate a common set of parameters.

In Figure 1, the open circles represent the fitted values (i.e., the predicted value of each case given the other data cases) based on the best fitted model; these fitted values appear to closely track the counts of serological positives. Figure 2 plots the standardized residuals against the fitted values. The plot suggests some outlying residuals, with the largest one equal to 3.14, which is not an outlier, after adjusting for multiple testing using Bonferroni inequality. This is because there are 76 data cases so that a residual is judged to be an outlier only if it exceeds 3.41 in magnitude, at 5% family error rate. Nonetheless, we have also refitted the best model reported above, but with the two largest "outliers" allowed for by incorporating dummy variables in the model; the estimates of θs become slightly smaller although their difference is still significant, that of τ is larger, and the rest of the parameter estimates essentially unchanged. Thus, we conclude that the model fit is robust to the presence of possible outliers.

In sum, we have demonstrated that (1) the immunity loss structure may be explained in terms of a 4-seasonal-lag disease structure, (2) the immunity loss rate is seasonal, but (3) the recovery rate is found to be non-seasonal.

[Table 1 about here.]

[Table 2 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

6. Conclusion

We have proposed a new stochastic framework useful for studying the dynamics of the immunity function of wildlife hosts against an infectious disease agent. The new proposed framework facilitates the study of the seasonal pattern of the immunity of great gerbils against plague. Using a common immunity structure, we illustrate this approach with the monitoring time-series data on the great gerbil system in Kazakhstan and show that the immune function of the great gerbil was seasonal. Interestingly, we find that the great gerbils in the monitoring study lost immunity to plague faster over the winter-to-summer season than over the summer-to-winter season. Earlier studies of viral, bacterial, and parasitic infections suggest that such seasonal pattern may be mediated by changes in food availability, temperature, photoperiod, and social behavioral changes; see Nelson *et al.* (2002), Klein *et al.* (2002) and Rogovin *et al.* (2003).

A major contribution of this paper consists of the development of an epidemiologically-based time series analysis of the immunity structure with aggregate time-series disease data. However, the continuous-time Markov chain framework may also be useful for analyzing panels of individual-based time-series disease data. Another interesting problem is how to extend our approach to include covariates so that we may identify important factors that may explain the observed seasonal structure.

Acknowledgements

This work was partially supported by the European Union as projects ISTC K-159 and STEPICA (INCO-COPERNICUS, ICA 2-CT2000-10048), as well as the authors' institutes. The work of SP was supported by the Post-doctoral Fellowship Program of Korea Science and Engineering Foundation (KOSEF). HV was supported by the Research Council of Norway (RCN).

Appendix A

Derivation of Equations (3) and (4)

We model the disease history of a (random) great gerbil by a continuous-time Markov chain with 4-states, namely, susceptible (S), infected (I), recovered (R) and dead (D). Denote by $\{X_t\}$ the Markov process whose state at time t can be either S, I, R or D. Each state has distinct implications on the outcomes of the bacteriological and/or serological test administered to the great gerbil. Specifically, if at time t a great gerbil has the state S, then it yields bacteriologically and serologically negative test outcomes, denoted by (B-,S-); if it is in state I, then it yields bacteriologically and serologically positive test outcomes, denoted by (B+,S+) (for simplicity, we have ignored the brief, initial period of infection when the great gerbil may not respond positively to either test); if it is in state R, then it yields bacteriologically negative and serologically positive test outcomes; denoted by (B-,S+).

The (infinitesimal) transition rate matrix of $\{X_t\}$ takes the following

form:

$$\begin{array}{c} \\ S \\ I \\ R \\ D \end{array} \begin{array}{c} \begin{array}{cccc} S & I & R & D \end{array} \\ \begin{pmatrix} -q_{SI}-q_{SD} & q_{SI} & 0 & q_{SD} \\ 0 & -q_{IR}-q_{ID} & q_{IR} & q_{ID} \\ q_{RS} & 0 & -q_{RS}-q_{RD} & q_{RD} \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

where for any two states x and y, the infinitesimal transition rate $q_{xy} = \lim_{\delta_t \downarrow 0} \frac{P(X_{t+\delta_t}=y|X_t=x)-P(X_t=y|X_t=x)}{\delta_t}$, and hence for small $\delta_t > 0$,
$q_{SI} \approx P(X_{t+\delta_t} = I|X_t = S)/\delta_t$, $q_{SD} \approx P(X_{t+\delta_t} = D|X_t = S)/\delta_t$, $q_{SS} = -q_{SI} - q_{SD} \approx \{P(X_{t+\delta_t} = S|X_t = S) - 1\}/\delta_t$,

$q_{IR} \approx P(X_{t+\delta_t} = R|X_t = I)/\delta_t$, $q_{ID} \approx P(X_{t+\delta_t} = D|X_t = I)/\delta_t$, $q_{II} = -q_{IR} - q_{ID} \approx \{P(X_{t+\delta_t} = I|X_t = I) - 1\}/\delta_t$,

$q_{RS} \approx P(X_{t+\delta_t} = S|X_t = R)/\delta_t$, $q_{RD} \approx P(X_{t+\delta_t} = D|X_t = R)/\delta_t$, $q_{RR} = -q_{RS} - q_{RD} \approx \{P(X_{t+\delta_t} = R|X_t = R) - 1\}/\delta_t$.

The Markov process has the following implications. If $X_t = x$ then the transition time to another state is exponential with mean $-1/q_{xx}$ (note q_{xx} is negative), and given the process has a transition, the (conditional) probability that the transition is from state x to state y is proportional to q_{xy}. Therefore, we can interpret q_{SI} as the (instantaneous) infective rate, $q_{IR}/(q_{IR}+q_{ID})$ as the recovery rate, $-1/q_{RR}$ as the mean immunity duration, $q_{ID}/(q_{IR}+q_{ID})$ as the mortality rate of plague and generally, $q_{SD} = q_{RD}$, as proportional to the natural death rates. In this application, we assume that q_{SI} is a free parameter that is constant within any given season. All other intensity parameters are, however assumed to be seasonal, i.e. taking one of two possible values, dependent on whether it is a winter-to-summer or summer-to-winter

season. (To simplify the derivation of the formulas below, all the intensity parameters will be treated as if they were constant.) The sampling was carried out in each season (spring and fall) at time Δ_t in the tth season. Plague prevalence was monitored by counting the number of bacteriologically and/or serologically positive great gerbils in the samples. First, we compute the (instantaneous) plague prevalence rate

$p_t = P(\text{B+ at } \Delta_t) = P(\text{a great gerbil tested bacteriologically positive at } \Delta_t)$

$= P(X_{\Delta_t} = I)$.

This probability can be approximately computed based on two working assumptions.

Assumption A: For each great gerbil, there is at most one successful plague attack within each season.

If recovering from a successful plague attack, the rodent may at any time be exposed to a new infectious flea bite. However, during 2-4 months after any acute period, a rodent possesses sufficient level of immunity to resist a new infection. Therefore, a great gerbil is not likely to be infected with plague more than once each season.

Assumption B: At the beginning of each season, the majority of the great gerbils are susceptible.

Recent field data suggests the validity of Assumption B at the beginning of the winter-to-summer seasons. Suleimenov *et al.* (2001) found that only about 10% of surviving great gerbils kept antibodies from fall to spring. Notice, however, that new cases of plague infections may appear early in spring. As antibodies may circulate for about 2 to 4 months and possibly longer (Kanatov, 1974), this assumption may be invalid for the summer-to-winter

season. During an acute epizootic, 50-70% of individuals in a given population may react serologically (Suleimenov *et al.*, 2001), and many of these individuals are likely to keep a high level of antibodies for several months. This suggests that the number of infected or recovered great gerbils could be about half of the number of susceptibles. Nonetheless, a careful analysis shows that the contributions to the probabilities of interest from the great gerbils that are infected or recovered are of smaller order of magnitude than those from the susceptibles. This is because at the beginning of tth season, (i) any infected great gerbil will not contribute to p_t by Assumption A, and (ii) any recovered great gerbil has to become susceptible before it can contribute to p_t, and hence such contribution is of smaller order of magnitude than contribution under Assumption B; similar reasoning applies to the computation of q_t. Consequently, Assumption B serves to simplify the ensuing calculations.

Based on these two working assumptions, the event $\{X_{\Delta_t} = I\}$ is almost equal to the event that there exists some intermediate time x between the beginning of the season, b_t, and the sampling time, Δ_t, such that the great gerbil is susceptible between b_t and x and becomes infected from x to Δ_t (and then perhaps some time thereafter). The probability of the later event

can be computed as follows.

$$\begin{aligned} p_t &= \int_{b_t}^{\Delta_t} (-q_{SS}) e^{q_{SS}(x-b_t)} q_{SI}/(-q_{SS}) P(\text{ the next transition time } \geq \Delta_t | X_x = I) dx \\ &= \int_{b_t}^{\Delta_t} q_{SI} e^{q_{SS}(x-b_t)} e^{q_{II}(\Delta_t - x)} dx \\ &= q_{SI} e^{q_{II}(\Delta_t - b_t)} \int_{b_t}^{\Delta_t} e^{(q_{SS}-q_{II})(x-b_t)} dx = q_{SI} e^{q_{II}(\Delta_t - b_t)} \frac{e^{(q_{SS}-q_{II})(x-b_t)}}{(q_{SS}-q_{II})} \Big|_{b_t}^{\Delta_t} \\ &= q_{SI} e^{q_{II}(\Delta_t - b_t)} \frac{e^{(q_{SS}-q_{II})(\Delta_t - b_t)} - 1}{(q_{SS}-q_{II})}. \end{aligned}$$

The immunity loss structure can be analyzed by studying the serological test data. To do this, we need to compute the following probability q_t=p($S+$ at Δ_t) = P(a great gerbil tested serologically positive at Δ_t)=$E(Q_t)/N_t$, where $E()$ denotes taking conditional expectation given the current and past rodent densities, N_t is the true rodent density in the tth season, Q_t is the number of great gerbils that test positive serologically at Δ_t and $Q_t = \sum_{j=0}^{M} Q_{t,j}$, where $Q_{t,j}$ is the number of great gerbils whose most recent infection occurr in the $(t-j)$th season and test positive serologically at Δ_t. The maximum values of M is set to be 7, because great gerbils rarely live beyond 4 years. Hence,

$$E(Q_t) = \sum_{j=0}^{M} N_{t-j} P(S+ \text{ at } \Delta_t, \text{most recent infection in the } (t-j)\text{th season}). \tag{A.1}$$

The probability P($S+$ at Δ_t, most recent infection in the $(t-j)$th season) can be computed as follows. Based on the two working assumptions A and B, the event of interest is almost equal to the event that a great gerbil is susceptible at b_{t-j}, the beginning of the $(t-j)$th season, then gets infected at x which is between b_{t-j} and L_{t-j}, the end of the $(t-j)$th season, and recovers at $x+y$

and remains so until Δ_t. Now, we introduce the third working assumption.

Assumption C: Any infection period is comparatively much shorter than any recovered period or susceptible period.

Generally, the infection period of great gerbils are less than 10 days, while antibodies to plague may be registered for several months (Kanatov, 1974).

Assumption C simplifies the calculation as we can take $y = 0$. Consider the case that $j > 0$. Then,

$$\begin{aligned}
& P(S+ \ \text{ at } \ \Delta_t, \\
& \quad \text{most recent infection in the (t}-\text{j)th season)} \\
\approx \ & \int_{b_{t-j}}^{L_{t-j}} (-q_{SS}) e^{q_{SS}(x-b_t)} q_{SI}/(-q_{SS}) \frac{q_{IR}}{(q_{IR}+q_{ID})} e^{q_{RR}(\Delta_t - x)} dx \\
= \ & e^{q_{RR}(\Delta_t - L_{t-j})} \int_{b_{t-j}}^{L_{t-j}} q_{SI} e^{q_{SS}(x-b_t)} e^{q_{RR}(L_{t-j}-x)} dx \times \frac{q_{IR}}{(q_{IR}+q_{ID})} \\
= \ & e^{q_{RR}(\Delta_t - L_{t-j})} P(\text{ a great gerbil infected in the (t}-\text{j)th season,} \\
& \quad \text{and keeps immunity throughout (t}-\text{j)th season).} \qquad \text{(A.2)}
\end{aligned}$$

Define the probability on the right side of the preceding equation by γ_{t-j}. Upon noting that the end of a season coincides with the beginning of the next season, we have

$$\begin{aligned}
e^{q_{RR}(\Delta_t - L_{t-j})} &= e^{q_{RR}(\Delta_t - b_t)} e^{q_{RR}(L_{t-1}-b_{t-1})} \cdots e^{q_{RR}(L_{t-j+1}-b_{t-j+1})} \\
&= e^{q_{RR}(\Delta_t - L_t)} e^{q_{RR}(L_t - b_t)} e^{q_{RR}(L_{t-1}-b_{t-1})} \cdots e^{q_{RR}(L_{t-j+1}-b_{t-j+1})} \\
&= e^{q_{RR}(\Delta_t - L_t)} \theta_t \cdots \theta_{t-j+1}, \qquad \text{(A.3)}
\end{aligned}$$

where $\theta_t = e^{q_{RR}(L_t - b_t)}$ which can be interpreted as the probability that a recovered great gerbil kept immunity throughout the tth season. Next, consider the case that $j = 0$, which means that the great gerbil is infected some

time within the tth season. We can compute the probability P($S+$ at Δ_t, most recent infection in the tth season) as follows:

$$\begin{aligned}
& \int_{b_t}^{\Delta_t} -q_{SS}e^{q_{SS}(x-b_t)}q_{SI}/(-q_{SS})\frac{q_{IR}}{(q_{IR}+q_{ID})}e^{q_{RR}(\Delta_t-b_t)}dx \\
= \; & q_{SI}e^{q_{RR}(\Delta_t-b_t)}\int_{b_t}^{\Delta_t} e^{(q_{SS}-q_{RR})(x-b_t)}dx \times \frac{q_{IR}}{(q_{IR}+q_{ID})} \\
= \; & q_{SI}e^{q_{RR}(\Delta_t-b_t)}\frac{e^{(q_{SS}-q_{RR})(x-b_t)}}{(q_{SS}-q_{RR})}\,|_{b_t}^{\Delta_t} \times \frac{q_{IR}}{(q_{IR}+q_{ID})} \\
= \; & q_{SI}e^{q_{RR}(\Delta_t-b_t)}\frac{(e^{(q_{SS}-q_{RR})(\Delta_t-b_t)}-1)}{q_{SS}-q_{RR}}\frac{q_{IR}}{(q_{IR}+q_{ID})} \\
= \; & \delta p_t\frac{(q_{SS}-q_{II})}{e^{q_{SS}(\Delta_t-b_t)}-e^{q_{II}(\Delta_t-b_t)}}\frac{e^{q_{SS}(\Delta_t-b_t)}-e^{q_{RR}(\Delta_t-b_t)}}{(q_{SS}-q_{RR})}\frac{q_{IR}}{(q_{IR}+q_{ID})}.
\end{aligned} \tag{A.4}$$

Let $\tau_t = \frac{(q_{SS}-q_{II})}{e^{q_{SS}(\Delta_t-b_t)}-e^{q_{II}(\Delta_t-b_t)}}\frac{e^{q_{SS}(\Delta_t-b_t)}-e^{q_{RR}(\Delta_t-b_t)}}{(q_{SS}-q_{RR})}\frac{q_{IR}}{(q_{IR}+q_{ID})}$. This expression can be simplified under the following assumption.

Assumption D: The magnitude of q_{SS} is much smaller than that of q_{RR} or q_{II}.

Assumption D amounts to the condition that the mean susceptible period $(-1/q_{SS})$ is much longer than the mean infection period $(-1/q_{II})$ or the mean recovered period $(-1/q_{RR})$, which seems to be consistent with the monitoring data under study (see the discussions below Assumptions (A-C)). Under this assumption and the natural condition that $\Delta_t - b_t$ is relatively large, $\tau_t \approx \frac{(q_{SS}-q_{II})}{(q_{SS}-q_{RR})}\frac{q_{IR}}{(q_{IR}+q_{ID})}$. It can be shown that γ_t has the same expression as τ_t except that Δ_t is replaced by L_t, and hence $\gamma_t \approx \tau_t$, under Assumption D. Combining (A.1)-(A.4), we get Equation (4).

References

Bhattacharya, R. N. and Waymire, E. C. (1990). *Stochastic processes with applications*. New York: John Wiley and Sons.

Cox, D. and Miller, H. (1968). *The Theory of Stochastic Processes*. Chapman and Hall, London.

Davidian, M and Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *J. Agr. Biol. Envir. Statist.*, **8**, 387-419.

Davis, S., Begon, M., De Bruyn, L., Ageyev, V. S., Klassovskiy, N. L., Pole, S. B., Viljugrein, H., Stenseth, N. Chr., and Leirs, H. (2004). Predictive Thresholds for Plague in Kazakhstan. *Science*, **304**, 736-8.

Dickmann, O. and Heesterbeek, J. A. P. (2000). *Mathematical Epidemiology of Infectious Disease: Model Building, Analysis and Interpretation*. Chichester: John Wiley.

Frigessi, A., Marshall, C., Holden, M., Viljugrein, H., Stenseth, N. Chr., Holden, L., Ageyev, V. and Klassovskiy, N. L. (2004). Bayesian Population Dynamics of Interacting Species: Great Gerbils and Fleas in Kazakhstan. To appear in *Biometrics*.

Grenfell, B. T. and Dobson, A. P. (1995). *Ecology of Infectious Diseases in Natural Populations*. New York: Cambridge University Press.

Kanatov, Yu. V. (1974). Systems of serological tests using sensitized erythrocytes for detecting antigens and antibodies. Thesis of doctoral dissertation. Saratov. 36 (in Russian).

Kanatov Yu.V., Kanatova E.A. & Misaleva O.S. (1969). *Systems of Serolog-*

ical Tests in Studies of Plague. II. Producing of Specific Antibodies in the Great Gerbils Being Experimentally Infected – Materials of the VI Scientific Conference of the Plague Control Institutions of Central Asia and Kazakhstan. Almaty, **1**, 31-33 (in Russian).

Klein S. L., Bird B. H., Nelson R. J. and Glass, G. E. (2002). Environmental and Physiological Factors Associated with Seoul Virus Infection Among Urban Populations of Norway Rats. *J. Mammal*, **83** 478-488.

Naumov, N. P. and Lobachev, V. S. (1975). Ecology of desert rodents of the U.S.S.R. (Jerboas and Gerbils). Great Gerbil. In *Rodents in desert environments* (eds. Prakash, I. and Ghosh, P.K), 549-598. The Hague: Dr. W. Junk b.v. Publishers.

Nelson, R. J., Demas, G. E., Klein, S. L. and Kriegsfeld, L. J. (2002). *Seasonal Patterns of Stress, Immune Function, and Disease.* New York: Cambridge University Press.

Perry, R. D. and Fetherston, J. D. (1997). *Yersinia pestis* – Etiologic Agent of Plague. *Clinical Microbiology Reviews*, **10**, 35-66.

Pinheiro, J. C. and Bates, D. M. (1995). *Mixed Effects Models in S and S-Plus.* New York: Springer.

Rogovin, K., Randall, J. A., Kolosova, I. and Moshkin, M. (2003). Social correlates of stress in adult males of the great gerbil, *Rhombomys opimus*, in years of high and low population densities. *Hormones and Behavior*, **43** 132-139.

Rothschild, V. E. (1978). *Spatial structure of plague natural focus and methods of its study.* Moscow: Publishing House of Moscow University. 192 (in Russian).

Suleimenov, B. M., Isin, Zh. M., Atshabar, B. B., Klassovskiy, N. L., Kogay, O. V. Kopbayev, E. Sh. and Novikov, G. S. (2001). Immunophysiological structure of acute plague epizootics. Quarantinable and Zoonotic Infections in Kazakhstan. Almaty, 4 261-266 (in Russian).

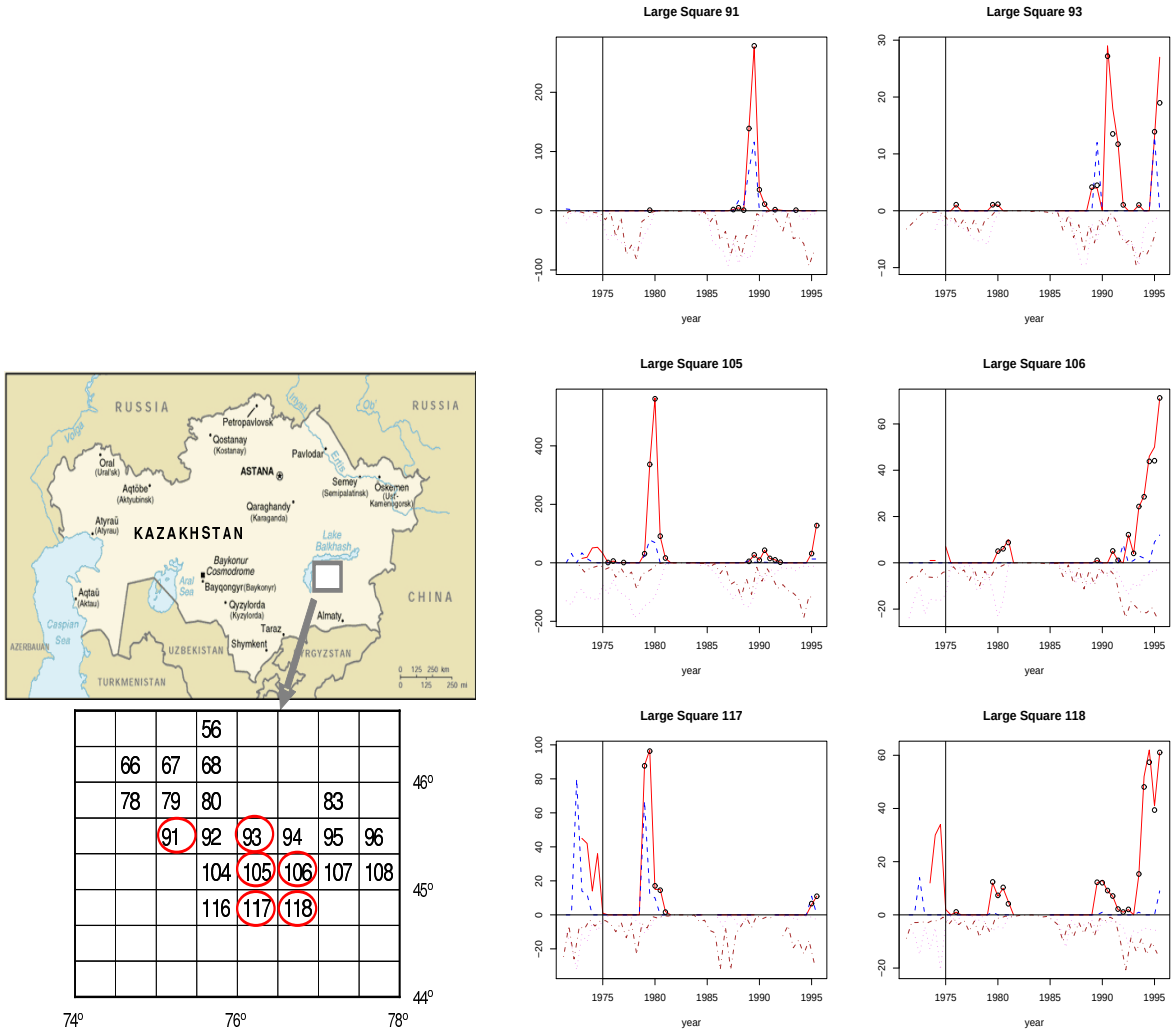

Figure 1. Upper left panel shows PreBalkhash focus marked as a square on a map of Kazakhstan. Lower left panel shows the large squares in the PreBalkhash focus from which we have data on the monitoring of plague prevalence. The right panel shows the time-series plots of counts of bacteriological positive (dashed line), serological positives (solid line), sample size (dot-dashed line) and rodent density (dotted line) for 6 large squares. The latter two variables are re-scaled and multiplied by a minus sign to render the time-series plots clearer. Open circles represent fitted values from a fitted model reported in Table 2.

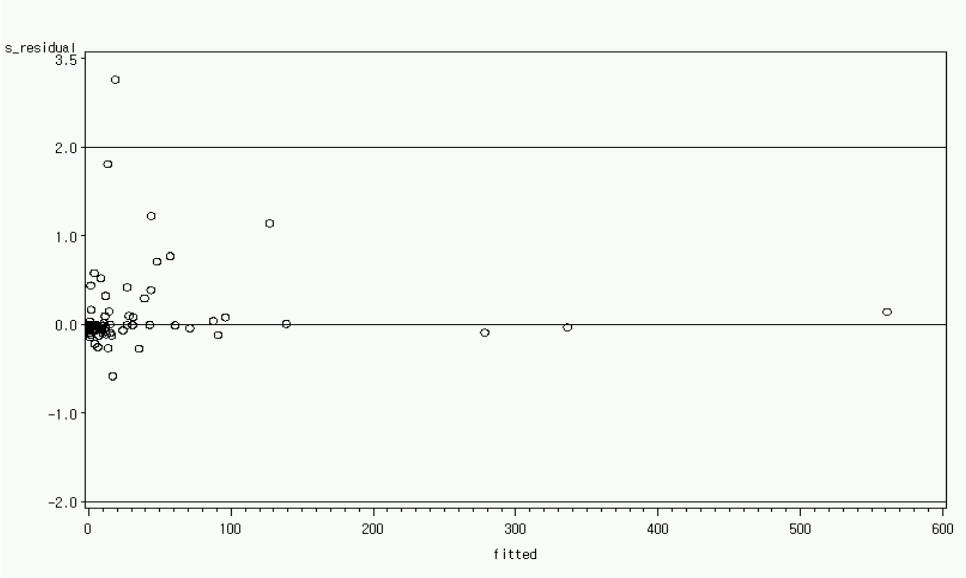

Figure 2. Standardized residual plot of the model reported in Table 2.

Table 1

AIC_C of the various common immunological models. Case 1: constant τ and seasonal θ, case 2: seasonal τ and constant θ, case 3: both τ and θ are constants and case 4: both τ and θ are seasonal.

	M							
Model	7	6	5	4	3	2	1	0
Case 1	35155.0	23007.0	1877.9	**752.6**	753.0	754.8	753.7	769.0
Case 2	35164.0	23017.0	1888.6	763.6	763.7	764.2	*765.4*	769.0
Case 3	*35168.0*	23021.0	1894.5	*769.6*	769.8	770.6	766.8	781.4
Case 4	35157.0	23009.0	1879.6	754.4	754.6	*756.3*	*757.0*	*771.7*

Table 2

Parameter estimates for the best model based on Table 1.

Parameter	Estimate	SE	P-value
θ_F	0.632	0.203	0.003
θ_S	0.135	0.030	<.0001
τ	7.256	2.598	0.007
σ^2	0.008	0.002	<.0001
$\theta_F - \theta_S$	0.497	0.204	0.017
C_{91}	1.5E-07	9.2E-05	0.999
C_{93}	0.073	0.029	0.014
C_{105}	0.032	0.024	0.180
C_{106}	0.074	0.027	0.007
C_{117}	0.013	0.046	0.776
C_{118}	0.096	0.023	<.0001