

Semiparametric estimation of the duration of immunity from infectious disease time series: influenza as a case-study

Yingcun Xia,

National University of Singapore, Singapore

Julia R. Gog

University of Cambridge, UK

and Bryan T. Grenfell

Pennsylvania State University, University Park, USA

[Received August 2003. Final revision August 2004]

Summary. An important epidemiological problem is to estimate the decay through time of immunity following infection. For this purpose, we propose a semiparametric time series epidemic model that is based on the mechanism of the susceptible–infected–recovered–susceptible system to analyse complex time series data. We develop an estimation method for the model. Simulations show that the approach proposed can capture the non-linearity of epidemics as well as estimate the decay of immunity. We apply our approach to influenza in France and the Netherlands and show a rapid decline in immunity following infection, which agrees with recent spatiotemporal analyses.

Keywords: Dynamical models in epidemics; Generalized partially linear single-index model; Immunity; Influenza; Kernel smoother; Susceptible–infected–recovered–susceptible model

1. Introduction

1.1. Preamble

For many viral and bacterial infections, the probability for a recovered (or vaccinated) host to be reinfected will increase as time following infection elapses. In other words, host immunity against disease will decay with time. Microparasitic infections are generally modelled in a compartmental framework, dividing a population between susceptible individuals (who are previously uninfected), infectious, recovered etc. groups (Anderson and May, 1991; Grenfell and Dobson, 1995). At one extreme, we have susceptible–infectious–recovered (SIR) models, where immunity is effectively lifelong. In fact, very few infections (notably measles and other morbillivirus infections of animals) fit this pattern (Grenfell *et al.*, 2001); for most diseases, individuals eventually lose immunity and can acquire infection to the same or (more likely) to a different strain of the pathogen (Andreasen *et al.*, 1997; Gupta *et al.*, 1994). The latter infections are often represented by susceptible–infectious–recovered–susceptible (SIRS) models (where individuals can have immunity) or susceptible–infectious–susceptible models, where

Address for correspondence: Yingcun Xia, Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, 117546 Singapore.
E-mail: staxyc@nus.edu.sg

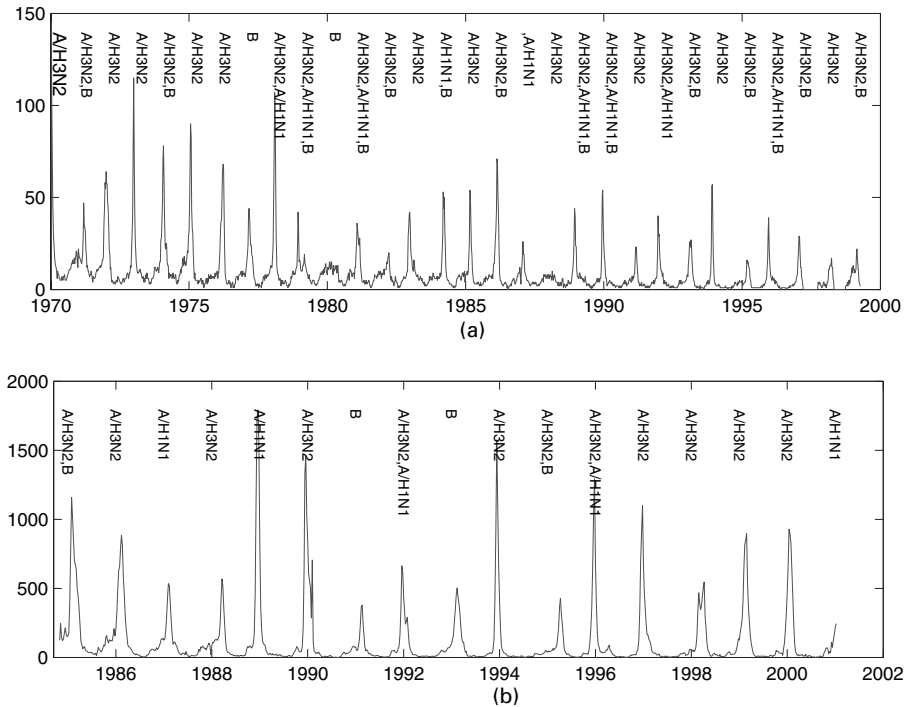


Fig. 1. (a) Weekly cases per 10000 in the Netherlands and (b) weekly cases per 100000 of influenza in France: the main types or strains for each year are labelled

immunity is very short (Hethcote and Yorke, 1984). For both practical and theoretical understanding of such complex strain dynamics, it is important to estimate the decay function of immunity for SIRS infections. Specifically, we consider the decay function of immunity after a host has recovered from the disease. We employ mechanistic time series models which capture the SIRS mechanism, extending previous work on the more straightforward SIR case (Finkenstädt and Grenfell, 2000).

We focus particularly on influenza. This is an acute viral infection of humans which is caused by a group of influenza virus A, B and C in decreasing order of importance (Nicholson *et al.*, 1998). Influenza is a very important human infection arising as a series of seasonal epidemics. Two notified time series of influenza-like illness cases in France and the Netherlands are shown in Fig. 1. In human influenza A, immunity to reinfection is finite, particularly because the virus undergoes a combination of year-to-year antigenic drift and occasional dramatic shift in haemagglutinin and neuraminidase surface protein; see Nicholson *et al.* (1998). These viral strain variations greatly complicate the design of vaccination strategies. We focus here on the dynamic effect of drift. Pease (1987) conjectured that immunity to influenza will decay linearly with time elapsed, whereas Couch and Kasel (1983) argued that immunity lasts for more than 4 years. Murphy and Clements (1989) investigated immunity for various age groups and showed significant differences between the groups. However, there have been few quantitative investigations of how immunity decays with time since infection. This problem is complicated by the presence of alternating or coexisting influenza subtypes, H2N2, H3N2 and H1N1, with partly distinct immunogenicity (Nicholson *et al.*, 1998; Ferguson *et al.*, 2003; Gog and Grenfell, 2002). The alternation of subtypes could greatly complicate the estimation of cross-immunity through time.

This is a particular potential problem in crude flu-like illness data (Fig. 1), which subsume all influenza A subtypes, the lower, more episodic incidence of influenza B and, potentially, other respiratory infections.

We first discuss the problem of capturing loss of immunity by indirectly estimating susceptibility dynamics. We then derive the model and the estimation method, testing it against model simulations. Finally, we fit the model to the influenza time series data, to estimate the immunity profile and the effect of subtype dynamics.

1.2. A semiparametric time series susceptible–infectious–recovered(–susceptible) model

Most SIR and SIRS models are based on a ‘mass action’ assumption that transmission is strictly proportional to the product of the densities of susceptible and infected individuals. Thus,

$$\frac{dz}{dt} \propto zS - cz, \quad (1.1)$$

where S is the density of susceptible individuals, z is the density of infected individuals and c is the recovery rate. Epidemic time series give quite a crude estimate of the number of infected individuals, e.g. measles data in England and Wales (which are reliable until the late 1980s; Grenfell and Harwood (1997)) or influenza-like illness data in France (which are rather more prey to uncertainty; see below).

In this paper, we first generalize model (1.1) by using semiparametric methods. The mass action assumption is a limitation of the model because there is increasing evidence that epidemic dynamics are more complicated than equation (1.1). Liu *et al.* (1987), Hochberg (1991) and Finkenstädt and Grenfell (2000) showed that adding exponents to S and z ,

$$\frac{dz}{dt} \propto z^\alpha S^\gamma - cz, \quad (1.2)$$

can lead to more complicated dynamics which can describe the observed data better in some cases (Finkenstädt and Grenfell, 2000). The mixing parameters of transmission, α and γ , give a phenomenological model for local spatial variation in the contact rate between the infectious hosts and susceptibles; these parameters may also compensate for the discretization that is inherent in the time series SIR approach (Glass *et al.*, 2003). From these considerations, model (1.2) is still far from reality. Instead of assuming possibly unrealistic functional forms as in expressions (1.1) or (1.2), we may allow the data themselves to decide the function. Therefore, we consider the flexible relationship

$$\frac{dz}{dt} \propto z^\alpha \nu(S) - cz, \quad (1.3)$$

where ν is any continuous function, which will be decided by the observed data. Our real data analysis below suggests that the generalization (1.3) is necessary. We call model (1.3) a semi-parametric mechanistic model. Similar approaches were introduced in Ellner *et al.* (1998).

To simplify the notation, we assume that the population size N is approximately constant. Thus model (1.3) still holds if we take z and S as the numbers of infected and susceptible individuals respectively. In practice, the notified data are discretized time series of newly infected cases y_t . According to the above discussion, we have $y_t \propto z_{t-1}^\alpha \nu(S_{t-1})$, where z_t is the number of infectious hosts in time period $t - 1$. If we choose the time unit to equal the infectious period, e.g. about 1 week for influenza, then z_{t-1} approximates the number of new infections in time period $t - 1$, i.e. $z_{t-1} \approx y_{t-1}$. We then have

$$y_t \propto y_{t-1}^\alpha \nu(S_{t-1}).$$

We call it the semiparametric time series SIR model or semiparametric time series SIRS model; see the discussion in Section 1.1.

2. The semiparametric time series susceptible–infectious–recovered–susceptible model

To use model (1.1), we must first reconstruct the number of susceptible individuals. Ellner *et al.* (1998) and Finkenstädt and Grenfell (2000) considered the case of non-recurring SIR diseases like measles. This is simpler than our current problem because, once a host is infected, under the SIR mechanism they will never be reinfected. Thus, the number of susceptible individuals is the total number in the population minus the cumulative sum of all previous cases reported. By contrast, for recurring infections, the number of susceptible individuals is a random variable. Those who have higher immunity against the disease have less probability of being infected or reinfected. Therefore, we shall estimate an expected number of susceptible individuals.

Let $p(i)$ denote the probability that a given host is susceptible in time period i after their last recovery from the disease. For simplicity, we make the following homogeneity assumption. The value of $p(i)$ depends only on the time i elapsed after the last recovery from the disease. Note that $p(i)$ is an increasing function. Suppose that p_0 is the limit of $p(i)$ as $i \rightarrow \infty$. We call $\kappa(i) = 1 - p(i)/p_0$ *relative immunity*. It is easy to see that $0 \leq \kappa(i) \leq 1$ and $\kappa(i)$ is a decreasing function of i . Suppose that there are H hosts and each has relative immunity $\kappa(i)$. Assume that the infections among these hosts are independent. By the notation above, the expected number of hosts that can be infected, i.e. the *expected number of susceptible individuals*, is $H p(i) = H p_0 \{1 - \kappa(i)\}$, where p_0 is defined above.

For the infected hosts y_{t-i} in time period $t-i$, by the above discussion they have probability $p(i)$ in time period t to be reinfected, i.e. their immunity in time period t is $\kappa(i)$. The expected number of susceptible individuals (within these y_{t-i} hosts) is $y_{t-i} p(i) = p_0 y_{t-i} \{1 - \kappa(i)\}$. If a person has never been infected before, we may take him equivalently as having been infected infinitely long ago. Thus the population $N = \sum_{i=1}^{\infty} y_{t-i}$. The total number of susceptible individuals in time period t is

$$\sum_{i=1}^{\infty} y_{t-i} p(i) = p_0 \left\{ N - \sum_{i=1}^{\infty} y_{t-i} \kappa(i) \right\}.$$

It is easy to see that $\kappa(i)$ will tend to 0 when i is sufficiently large. Let m be the period in which a recovered host has immunity (i.e. $\kappa(i) = 0$ if $i > m$). If the immunity can last for a whole life, then m is the lifespan. If the population size is known, on the basis of the discussion in Section 1 we may assume that

$$y_t = \beta_t y_{t-1}^\alpha \nu \left[p_0 \left\{ N - \sum_{i=1}^m y_{t-i} \kappa(i) \right\} \right].$$

Considering random effects and the unknown population size, we finally propose the following semiparametric time series SIRS model:

$$E\{y_t | y_{t-i}, i \geq 1\} = \beta_t y_{t-1}^\alpha \mu \left\{ \sum_{i=1}^m \kappa(i) y_{t-i} \right\} \quad \text{subject to } \kappa(i) \geq \kappa(i+1), i = 1, 2, \dots, m-1, \quad (2.1)$$

where $\mu(x) = \nu\{p_0(N-x)\}$, which describes the functional relationship between the expected number of immune hosts and expected cases in the next time unit, and

$$\beta_t = \exp\left(\sum_{k=1}^{\tau} \rho_k D_{k,t}\right).$$

Here, $\rho_k, k=1, 2, \dots, \tau$, are parameters for seasonal forces, τ is the period of an epidemic cycle and $D_{k,t}$ is employed to describe seasonal variations in infection rate: $D_{k,t} = 1$ if $k = t(\text{mod}, \tau)$; $D_{k,t} = 0$ otherwise.

If the under-reporting rate is approximately constant, model (2.1) can still be used. Suppose that the under-reporting rate is r ; then the link function $\mu\{\sum_{i=1}^m \kappa(i)y_{t-i}\}$ in model (2.1) becomes $\mu\{\sum_{i=1}^m \kappa(i)y_{t-i}/r\}$. Since $\mu(\cdot)$ is unspecified, we can redefine $\mu(\cdot)$ as $\mu(\cdot/r)$.

Following Finkenstädt and Grenfell (2000), we consider the stochastic model

$$y_t = \beta_t y_{t-1}^\alpha \mu\left\{\sum_{i=1}^m \kappa(i)y_{t-i}\right\} \xi_t, \quad (2.2)$$

where $E\{\log(\xi_t)|y_{t-i}, i \geq 0\} = 0$ almost surely. Taking logarithms of both sides of equation (2.2), we have

$$z_t = U_t^T \beta + g(X_t^T \theta) + \varepsilon_t, \quad \text{subject to } \theta_1 \geq \theta_2 \dots \geq \theta_m \geq 0. \quad (2.3)$$

Here, $g = \log(\mu)$, $\beta = (\rho_1, \rho_2, \dots, \rho_\tau, \alpha)^T$ and $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T = (\kappa(1), \dots, \kappa(m))^T$, $U_t = (D_{1,t}, D_{2,t}, \dots, D_{\tau,t}, z_{t-1})^T$, $X_t = (y_{t-1}, y_{t-2}, \dots, y_{t-m})^T$, $z_t = \log(y_t)$ and $\varepsilon_t = \log(\xi_t)$. The first part of equation (2.3) is the generalized partially linear single-index model that was proposed by Carroll *et al.* (1997); see also Xia and Härdle (2002) under a time series setting. Note that, in the model, g is an unknown function and β and θ are unknown parameters. Model (2.2) is based on epidemiological knowledge and therefore has some prior information. This prior knowledge leads to some constraints on the model. Existing estimation methods are not suitable for us, so we shall propose a new estimation method for the constrained model.

3. Estimation method

Now, our main focus is on how to estimate the parameters $\theta_i = \kappa(i)$, $i=1, \dots, m$, given observed values for X and U . Because both g and (β, θ) are unknown, we shall first approximate g by local linear functions. Thus we estimate g by solving a local linear regression. We shall then estimate (β, θ) globally on the basis of estimated g . To make it identifiable, we further assume that $\kappa(1) = \theta_1 = 1$. Our interest is the relative immunity, and the assumption does not change the relationship between the $\kappa(i)$, $i=1, 2, \dots, m$, under the model setting. We assume that (X_t, U_t, z_t) , $t=1, 2, \dots, n$, from model (2.2), have the same marginal distribution as (X, U, z) . More precisely, our estimation method can be stated as follows. Given the link function g , the true parameter vectors θ and β , say θ_0 and β_0 , minimize

$$E[z - \{\beta^T U + g(\theta^T X)\}]^2. \quad (3.1)$$

The conditional variance given X , U and (β, θ) is

$$\sigma_{\beta, \theta}^2(\theta^T X, \beta^T U) = E\{[z - \{\beta^T U + g(\theta^T X)\}]^2 | \theta^T X, \beta^T U\}.$$

It follows that

$$E[z - \{\beta^T U + g(\theta^T X)\}]^2 = E\{\sigma_{\beta, \theta}^2(\theta^T X, \beta^T U)\}.$$

Therefore, minimizing expression (3.1) is equivalent to minimizing, with respect to θ and β ,

$$E\{\sigma_{\beta,\theta}^2(\theta^T X)\}. \quad (3.2)$$

Let $\{y_t, t = 1, 2, \dots, T\}$ be a realization from model (2.2). Because we use its lag variables for $X_t = (y_{t-1}, \dots, y_{t-m})$, we can use only $\{(X_t, U_t, z_t) : t = m+1, 2, \dots, T\}$ to estimate the model. Let $n = T - m$. For any given X_0 , a local linear expansion of $g(\theta^T X_i)$ at $\theta^T X_0$ is

$$g(\theta^T X_i) = g(\theta^T X_0) + g'(\theta^T X_0)\theta^T X_{i0} + O_P\{(\theta^T X_{i0})^2\},$$

where $X_{i0} = X_i - X_0$. Hence, we have the following approximation for X_i close to X_0 :

$$z_i - \beta^T U_i - g(\theta^T X_i) \approx z_i - \beta^T U_i - g(\theta^T X_0) - g'(\theta^T X_0)X_{i0}^T \theta.$$

Following the idea of local linear kernel estimation (Fan and Gijbels, 1996), we may estimate $\sigma_{\beta,\theta}^2(\theta^T X_0)$ by

$$\hat{\sigma}_{\beta,\theta}^2(\theta^T X_0) = \min_{a,d} \left\{ \sum_{i=1}^n (z_i - \beta^T U_i - a - dX_{i0}^T \theta)^2 w_{i0} \right\}. \quad (3.3)$$

Here, $w_{i0} \geq 0$, $i = 1, 2, \dots, n$, are some weights with $\sum_{i=1}^n w_{i0} = 1$, typically centred at X_0 . Let $X_{ij} = X_i - X_j$. By expressions (3.2) and (3.3), our estimation procedure is to minimize

$$n^{-1} \sum_{j=1}^n \rho(X_j, U_j) \sum_{i=1}^n (z_i - \beta^T U_i - a_j - d_j X_{ij}^T \theta)^2 w_{ij} \quad (3.4)$$

with respect to (a_j, d_j) and (β, θ) , where ρ is a weight function; see Appendix A for more details. The implementation of minimizing expression (3.4) and selection of pilot parameters are also discussed in Appendix A.

The time period m in model (2.2) is usually very large. For example, $m = 260$ if we consider weekly immunity for 5 years. Together with the other parameters, 52 for yearly seasonality and α , there are 313 parameters. With so many parameters, the estimation is usually quite unstable. In our case, we have constraints on the parameters. Our simulations suggest that using these constraints can reduce the variation of estimation substantially. See also the discussion in Liew (1976). Another possible way to reduce the number of parameters and the variation in estimation is to ignore the small difference of immunity in certain adjacent weeks and to assume that they are the same. However, as the number of parameters reduces, the estimation bias will increase. There is a trade-off between stability and the bias. Theory for choosing the number of parameters needs intensive investigation.

4. Simulations

To check the performance of our estimation method for finite data sets, we apply the method to data sets that were generated from the discrete SIRS model

$$E(y_t | y_{t-i}, i > 1) = \beta_t y_{t-1}^\alpha \left\{ N - \sum_{i=1}^{104} \kappa(i) y_{t-i} \right\}^\gamma, \quad (4.1)$$

where $\beta_t = [0.4 \sin\{2\pi t(\text{mod}, 52)/52\} + 2]/N^\gamma$ with $N = 500000$, $\alpha = 1$ and $\gamma = 0.9$. Model (4.1) has annual (52 weeks) seasonality. When $i \leq 104$, the decay functions of immunity are set to be $\kappa(i) = 1 - i/104$ and $\kappa(i) = \{1 - (i - 52)^3/52^3\}/2$. When $i > 104$, they are 0. The immunity functions are shown in Fig. 2. We assume that $\log(\xi_t) = 0.1\varepsilon_t$ and

$$\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, 1).$$

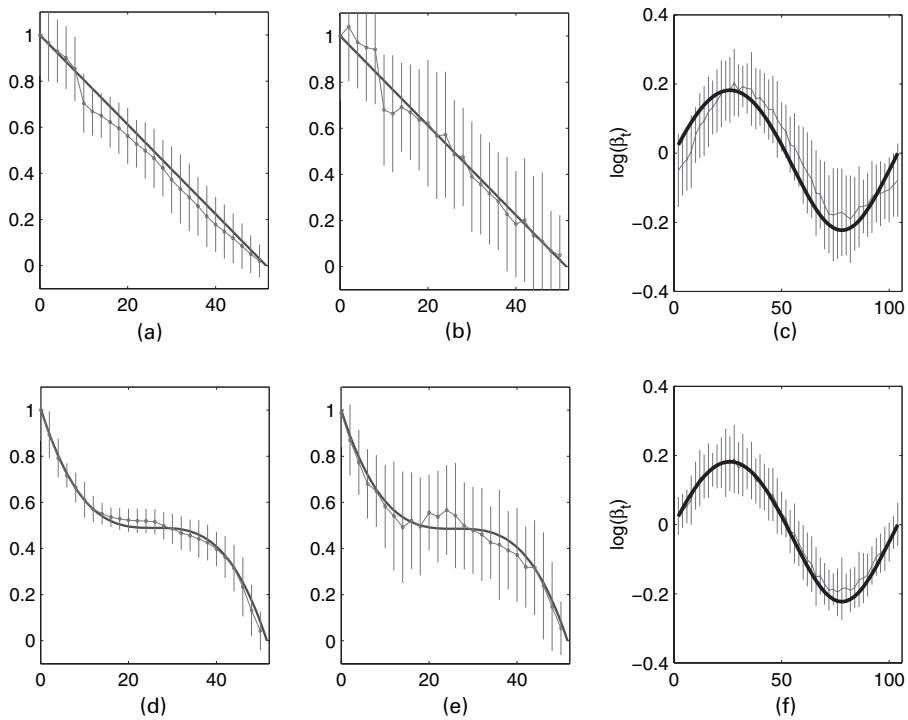


Fig. 2. (a) Decay function estimates with constraints, (b) decay function estimates without constraints, (c) estimates of seasonal transmission parameters (with a shift), (d) decay function estimates with constraints, (e) decay function estimates without constraints and (f) estimates of seasonal transmission parameters (with a shift): in (a), (b), (d) and (e) the bold curves are the true decay functions of immunity and the light curves are the averages of the estimated decay functions (\bar{f} , two times the standard deviations of the estimates)

As discussed in Section 3, we need to reduce the number of parameters for the immunity m in the estimation. In the following simulations, we assign the same value to parameters of four adjacent weeks. With sample size 520 (10 years), 100 independent realizations are drawn from model (4.1). The estimation results are shown in Figs 2(a) and 2(d). The method proposed can estimate the decay function of immunity quite accurately. However, there is a bias when the true decay function has big curvature, i.e. the second derivative is far from 0. This might be the result of reducing the number of parameters. The means (and standard errors) of the estimates of parameter α based on data under different settings corresponding to Figs 2(a) and 2(d) are 0.9913 (0.0059) and 0.9924 (0.0241) respectively. The estimators of seasonality parameters are also accurate and stable. See Figs 2(c) and 2(f). The unknown function $g(\cdot)$, i.e. the relationship between the expected number of immune people and the number of new cases, can also be estimated accurately. As a comparison, we also estimate the decay function of immunity without using constraint $\theta_1 \geq \theta_2 \geq \dots \geq \theta_m \geq 0$. The estimation results are shown in Figs 2(b) and 2(e). The estimations without constraints are not so stable as those with constraints.

5. Analysis of influenza

We now apply our semiparametric time series SIRS model to influenza data in France and the Netherlands (courtesy of Dr A. Bartelds and Dr P. Spreeuwenberg of the Netherlands Institute of Primary Health Care and Dr Cécile Viboud); see Fig. 1. The data comprise crude numbers of

people who were hospitalized with flu-like illness and therefore may subsume other respiratory infections (Nicholson *et al.*, 1998).

For both the French and the Netherlands data, we were also kindly given qualitative data on the alternation of subtypes (whether the subtypes were present in a given year). Initially, we choose $m = 400$ weeks and assign the same immunity to adjacent 4-week periods to reduce the number of parameters. The estimation result suggests that the immunity is zero after about 250 weeks in France and 350 weeks in the Netherlands. To refine the estimates, we take $m = 300$ for France and $m = 380$ for the Netherlands. We obtained the following estimation results.

For the data from France, we estimate that $\hat{\alpha} = 0.94$ (SE = 0.012) and that the variance of $\varepsilon_t (= \log(\xi_t))$ is 0.126. Note that the variance of $z_t = \log(y_t)$ is 2.2608. The proportion of the variance of $\{z_t\}$ that can be explained by the model is $R^2 = 94.4\%$. We explore a lack-of-fit test for the model by using Bartlett's Kolmogorov–Smirnov statistic; see, for example, Tong (1990). By checking the autocorrelation coefficients up to order 500, the test statistic is 0.4876, which suggests that the model fits the data quite well. For the data from the Netherlands, we have that $\hat{\alpha} = 0.90$ (SE = 0.020) and that the estimated variance of ε_t is 0.1025 and the variance of z_t is 0.8406. The proportion of the variance of z_t that can be explained by the model is $R^2 = 87.8\%$. By checking the autocorrelation coefficients up to order 500, Bartlett's Kolmogorov–Smirnov test statistic is 0.1191, which also suggests that the model can be accepted.

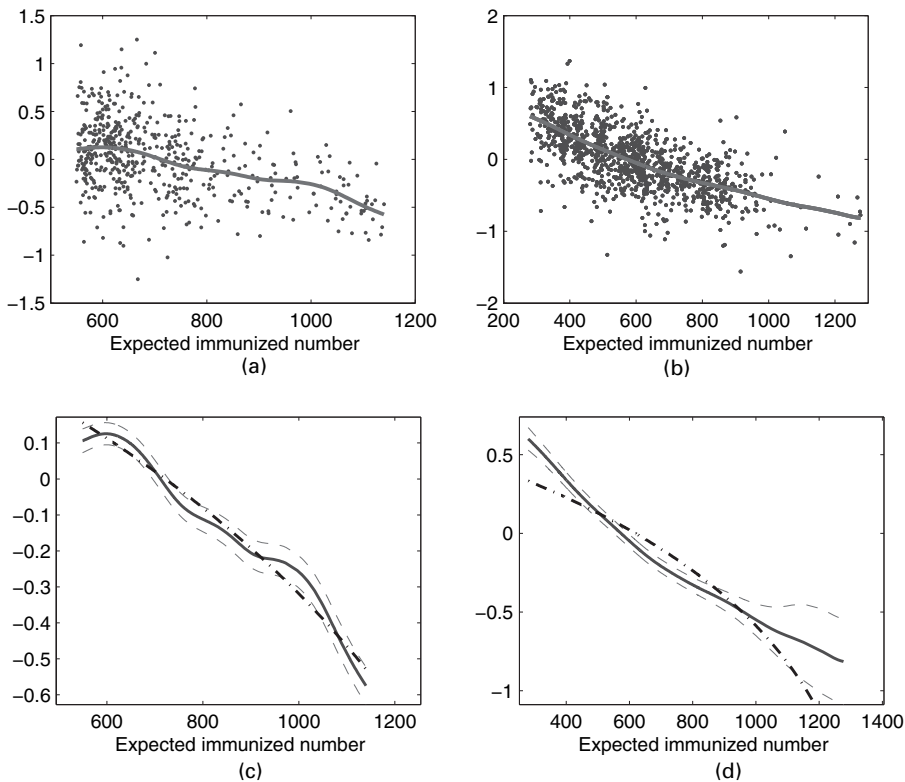


Fig. 3. Estimated functions of $g = \log(\mu)$ (—) for (a) France and (b) the Netherlands (•, corresponding residuals after removing the linear part in model (2.3); the x -axis is the expected immunized number $s = \sum_{i=1}^m \kappa(i) y_{t-i}$ per 10000 population) and (c), (d) 95% confidence intervals for g (—) and $g_0(s) = a + b \log(N - s)$ that minimizes the difference between g_0 and g with respect to a , b and N (---) (note that these g_0 s are not fully in the confidence intervals)

The estimated functions g are shown in Figs 3(a) and 3(b) for France and the Netherlands respectively. Although the epidemics in France and the Netherlands have different forms as shown in Fig. 1, it is interesting to see that the basic signature of immunity appears the same.

- (a) The expected number of immune hosts has a significant negative effect on the number of cases in the next time unit. The correlation between the number of people with immunity $\sum_{i=1}^m \kappa(i) y_{t-i}$ and $\log(y_t) - \beta^T U_t$, i.e. $\log(y_t)$ with the effect of other factors (seasonal variation and y_{t-1}) removed, in Figs 3(a) and 3(b) are -0.4134 and -0.6696 respectively for France and the Netherlands (significant at $p = 0.001$). This is in line with the SIRS mechanism. However, the quantitative relationship between y_t , y_{t-1} and the expected number of susceptible individuals, as shown in Figs 3(a) and 3(b), is more complicated than ‘mass action’ as expected in model (1.1) and its generalization (1.2). See Figs 3(c) and 3(d).
- (b) The epidemics in France and the Netherlands have similar decay functions of immunity as shown in Figs 4(a) and 4(b).
- (c) The estimated seasonal infection rates (the points in Figs 4(c) and 4(d)) agree with the general observation that, in the winter, the forces of infection are stronger than at other periods (Nicholson *et al.*, 1998).

Note that the decay pattern of immunity is different from the conjecture of Pease (1987). In the first 2 months, the recovered hosts have a high level of immunity. After that, the immunity decreases quite quickly. After about 8–12 months, the immunity level is relatively low. However, this low level of immunity will last for as long as another 4 or 5 years. (A statistical test that was developed in Xia and Härdle (2002) suggests that they differ from 0 in the first 3 years after recovering from the influenza at significance level 0.05). After 6 years, the immunity has been lost totally. Considering explanations for the decay of immunity, two factors emerge as follows.

- (a) The fast decay of immunity at the beginning may partly reflect hospital notification biases—if individuals are rapidly reinfected, they may not have clinical infections. It might also reflect the fact that the speed of drift could in turn depend on the number of cases; see Boni *et al.* (2004).
- (b) The subsequent slow decay pattern could reflect gradual effects of drift, though this is a complex picture because ‘flu-like’ illness subsumes influenza B, influenza A subtypes and possibly other respiratory infections. The overall immunity can last as long as 5 years though it is relatively weak. In this sense, our conclusion agrees with the arguments of Couch and Kasel (1983) and Murphy and Clements (1989).

Finally, Fig. 5 explains whether there is a signature of subtype dynamics (also subsuming flu type B incidence) in the residuals from our model. No clear influence is apparent: this might reflect the crudeness of the data, or—more interestingly—indicate that short-term immunity is ‘strain transcending’ (Ferguson *et al.*, 2003).

6. Conclusions

In this paper, we proposed a semiparametric SIRS model to estimate the decay of immunity for recurring diseases. Both simulation and analysis of empirical data indicate that the model can estimate the decay of immunity efficiently and can also capture other basic features of the epidemics.

The motivation of the semiparametric model is based on the observation that the SIR model lacks flexibility; see for example Liu *et al.* (1987) and Finkenstädt and Grenfell (2000). Note

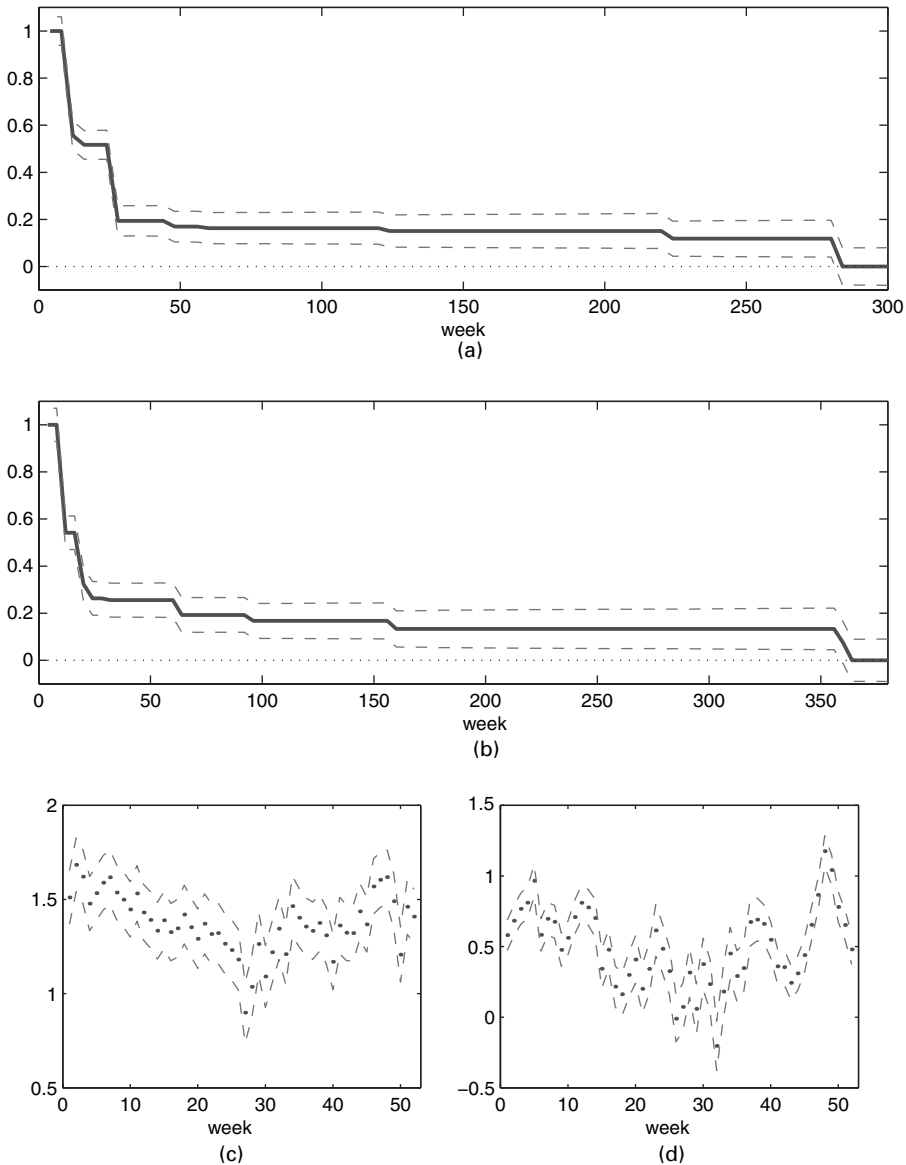


Fig. 4. Estimated decay functions of immunity for (a) France and (b) the Netherlands (— — —, 95% confidence interval) and estimated seasonal forces for (c) the Netherlands and (d) France (— — —, 95% confidence intervals)

that the estimates of the function $g(x)$ in Fig. 3 are different from the parametric assumption $g(v) = \log(N - v)$. Statistical tests in Figs 3(c)–3(d) suggest that the difference is significant; see Xia and Härdle (2002) for details. The violation of the parametric assumption might result from local spatial structure in transmission—this is an important area for future work. Further statistical generalizations of the model are possible. For example, we can allow the relationship of y_t and y_{t-1} to be more flexible and the mass action to become

$$y_t \propto \lambda(y_{t-1}) \nu(S_{t-1}),$$

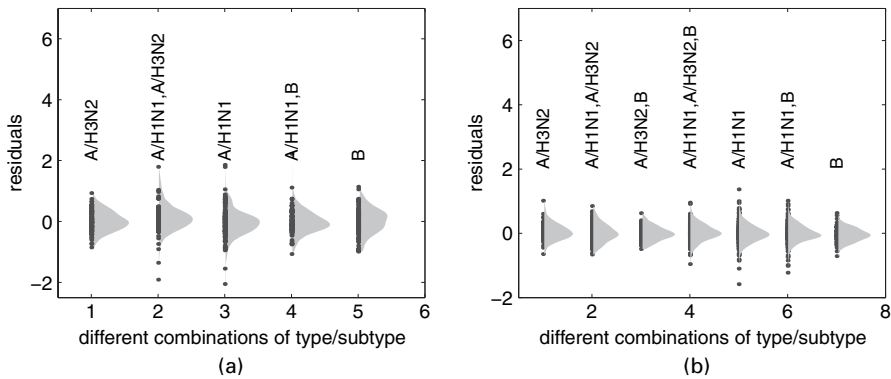


Fig. 5. Fitted residuals (●) plotted against various combinations of influenza type and subtypes for (a) France and (b) the Netherlands: the subtype and influenza type B are presented as presence and absence estimates for a given year (■, corresponding density of the residuals)

where λ is an unknown function. Corresponding to expression (2.3), we obtain a more general model

$$z_t = \beta^T U_{1t} + \phi(z_{t-1}) + g(\theta^T X_t) + \varepsilon_t,$$

where both $\phi = \log(\lambda)$ and $g = \log(\nu)$ are unknown functions, $z_t = \log(y_{t-1})$, $U_{1t} = (D_{1,t}, D_{2,t}, \dots, D_{\tau,t})^T$ and $\beta_1 = (\rho_1, \rho_2, \dots, \rho_\tau)^T$. This model is attractive from a statistical point of view and can be estimated by following our approach. However, our limited real data sets suggest that this generalization is not necessary.

The main epidemiological result is the strongly biphasic nature of the influenza immunity function (Figs 4(a) and 4(b)): a high level of initial immunity rapidly declines over a few months, followed by a much slower decay of the remaining immunity. Interestingly, the rapidly decaying component matches a recent proposal by Ferguson *et al.* (2003), that short-term cross-strain immunity may be a crucial feature of the population dynamics and resultant phylogenetics of influenza. The fact that there is no clear subtype-specific signature in the short-term immunity function is consistent with the hypothesis that it is strain transcending.

As discussed above, our analysis is based on data from influenza-like illness. This is a crude measure of overall levels of influenza and is probably biased towards older people. Age-structured laboratory-confirmed cases for influenza or other infections will provide much better material for future studies of the approach. It would be of significant value to include the effects of covariates giving prevalence of different subtypes in circulation. However, our limited analysis indicates that residual patterns from the correct fit do not covary with crude overall estimates of the relative frequency of strains H3N2 and H1N1 in laboratory samples. Also, as more data become available from mass screening, a key extension of this work is to incorporate the geography of the host.

Acknowledgements

The Joint Editor and three reviewers provided valuable comments on the paper. This research was supported by the Wellcome Trust (BG and YX), the Academic Research Fund of the National University of Singapore, the Biotechnology and Biological Science Research Council (BG) and Queens College, Cambridge (JRG). We thank Dr A. Bartelds and Dr P. Spreeuwenberg of the Netherlands Institute of Primary Health Care and Dr Cécile Viboud of the Fogarty

International Center, US National Institutes of Health, for providing the influenza data and useful discussions.

Appendix A: Implementation of the estimation and asymptotic distributions

The estimation procedure in Section 3 can be implemented by using any semiparametric regression methods. Here, we adopt the popular nonparametric local linear kernel smoothing method. The smoothing method approximates the unknown link function such as g in our model by a linear function locally and estimates the function by a weighted least squares estimation method. The weights are calculated by using a probability density function, called a kernel function; see Fan and Gijbels (1996). The details of our estimation methods are as follows. Given θ and β , we have

$$\begin{pmatrix} a_j \\ d_j \end{pmatrix} = \left\{ \sum_{i=1}^n w_{ij} \begin{pmatrix} 1 \\ X_{ij}^T \theta \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij}^T \theta \end{pmatrix}^T \right\}^{-1} \sum_{i=1}^n w_{ij} \begin{pmatrix} 1 \\ X_{ij}^T \theta \end{pmatrix} (z_i - \beta^T U_i). \quad (\text{A.1})$$

Given a_j and d_j , $j = 1, 2, \dots, n$, to calculate θ and β , we minimize

$$(\beta^T, \theta^T) D \begin{pmatrix} \beta \\ \theta \end{pmatrix} - 2c^T \begin{pmatrix} \beta \\ \theta \end{pmatrix} \quad \text{subject to } 1 \geq \theta_2 \geq \dots \geq \theta_m \geq 0, \quad (\text{A.2})$$

where

$$D = n^{-1} \sum_{j=1}^n \rho(X_j, U_j) \sum_{i=1}^n w_{ij} \begin{pmatrix} U_i \\ X_{ij} d_j \end{pmatrix} \begin{pmatrix} U_i \\ X_{ij} d_j \end{pmatrix}^T, \\ c = n^{-1} \sum_{j=1}^n \rho(X_j, U_j) \sum_{i=1}^n w_{ij} \begin{pmatrix} U_i \\ X_{ij} d_j \end{pmatrix} (z_i - a_j).$$

The quadratic programming problem in expression (A.2) can be solved by using quite standard algorithms; see, for example, Fletcher (1987).

The choice of the kernel function is as follows. Suppose that $H(\cdot)$ and $K(\cdot)$ are an m -variate and a univariate density function respectively. We first use weights

$$w_{ij,b} = H_{b,i}(X_j) / \sum_{l=1}^n H_{b,l}(X_j),$$

where $H_{b,i}(x) = b^{-m} H\{(X_i - x)/b\}$ and b is a bandwidth. This is a multivariate dimensional kernel weight. Because of the so-called ‘curse of dimensionality’ in nonparametrics, the estimation based on this kind of weight is not efficient. However, the multivariate kernel can help us to find an appropriate initial value of the estimate. We then use weights

$$w_{\gamma,ij,h} = K_{\gamma,h,i}(\gamma^T X_j) / \sum_{l=1}^n K_{\gamma,h,l}(\gamma^T X_j),$$

where $K_{\gamma,h,i}(v) = h^{-1} K\{(\gamma^T X_i - v)/h\}$, h is the bandwidth and γ is the previous estimate of θ . This is a single-index kernel weight. The minimization in expression (3.4) can be solved by iterations between expressions (A.1) and (A.2). Denote the estimators of θ and β by $\hat{\theta}$ and $\hat{\beta}$ respectively. After obtaining estimates $\hat{\theta}$ and $\hat{\beta}$, we can then estimate $g(v)$ by the solution of a_j in equation (A.1) with $X_j^T \theta$ (and $X_j^T \gamma$) replaced by v . The consistency of the estimators was proved in Xia and Härdle (2002).

We introduce the weight function ρ to handle the boundary points for technical purposes as follows. For any θ , we can have an estimate for the density of $\theta^T X$ as

$$\hat{f}_\theta(v) = (nh)^{-1} \sum_{i=1}^n K\{(\theta^T X_i - v)/h\}.$$

We define $\rho(v) = \hat{f}_\theta(v)$ or $\rho(v) = 1$ if $\hat{f}_\theta(v) > c_0$, and $\rho(v) = 0$ otherwise, where c_0 is a constant. If we assume that $\theta_0^T X$ has compact positive support, we can further take $\rho \equiv 1$. In our calculations, $\rho \equiv 1$ is used.

An advantage of the estimation method above is that undersmoothing is unnecessary for the estimators of parameters to achieve root n consistency; see Xia and Härdle (2002). Thus, most bandwidth selection

methods can be employed here. Next, we give the details for the cross-validation bandwidth selection method. Given θ and β , let

$$\begin{aligned} \text{CV}_0(b) &= n^{-1} \sum_{i=1}^n (z_i - \beta^T U_i - \tilde{a}_i^{\setminus i})^2, \\ \text{CV}(h) &= n^{-1} \sum_{i=1}^n (z_i - \beta^T U_i - \hat{a}_i^{\setminus i})^2, \end{aligned}$$

where $\tilde{a}_i^{\setminus i}$ and $\hat{a}_i^{\setminus i}$ are respectively given in

$$\begin{pmatrix} \tilde{a}_i^{\setminus i} \\ \tilde{d}_i^{\setminus i} \end{pmatrix} = \left\{ \sum_{\substack{j=1 \\ j \neq i}}^n w_{ji,b} \begin{pmatrix} 1 \\ X_{ji}^T \theta \end{pmatrix} \begin{pmatrix} 1 \\ X_{ji}^T \theta \end{pmatrix}^T \right\}^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n w_{ji,b} \begin{pmatrix} 1 \\ X_{ji}^T \theta \end{pmatrix} (z_j - \beta^T U_j)$$

and

$$\begin{pmatrix} \hat{a}_i^{\setminus i} \\ \hat{d}_i^{\setminus i} \end{pmatrix} = \left\{ \sum_{\substack{j=1 \\ j \neq i}}^n w_{\theta,ji,h} \begin{pmatrix} 1 \\ X_{ji}^T \theta \end{pmatrix} \begin{pmatrix} 1 \\ X_{ji}^T \theta \end{pmatrix}^T \right\}^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n w_{\theta,ji,h} \begin{pmatrix} 1 \\ X_{ji}^T \theta \end{pmatrix} (z_j - \beta^T U_j).$$

The bandwidth in each step is then $\hat{b} = \arg \min_b \{\text{CV}_0(b)\}$ for the weight function $w_{i,j,b}$ and $\hat{h} = \arg \min_h \{\text{CV}(h)\}$ for the weight function $w_{\theta,ji,h}$.

Let $\mu(x|\theta) = E(X|\theta^T X = \theta^T x)$, $\nu(x|\theta) = E(U|\theta^T X = \theta^T x)$ and

$$W_0 = E \left\{ \begin{pmatrix} U - \nu(X|\theta_0) \\ g'(\theta_0^T X) \{X - \mu(X|\theta_0)\} \end{pmatrix} \begin{pmatrix} U - \nu(X|\theta_0) \\ g'(\theta_0^T X) \{X - \mu(X|\theta_0)\} \end{pmatrix}^T \right\}.$$

Under some conditions, we have the following limiting distribution for the estimators:

$$n^{1/2} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{D} N(0, W_0^{-1} \sigma^2),$$

and

$$(nh)^{1/2} \left\{ \hat{g}(v) - g(v) - \frac{1}{2} g''(v) h^2 \right\} \xrightarrow{D} N \left\{ 0, f^{-1}(v) \int K^2(v) dv \sigma^2 \right\},$$

where $\sigma^2 = \text{var}(\varepsilon_i)$ and $f(v)$ is the density function of $\theta_0^T X$. If $h = o(n^{-1/5})$, then the 95% confidence interval for $g(v)$ is

$$\left[\hat{g}(v) - 1.96 f^{-1}(v) \int K^2(v) dv \sigma^2, \hat{g}(v) + 1.96 f^{-1}(v) \int K^2(v) dv \sigma^2 \right].$$

All the terms above can again be estimated by using the kernel smoothing method.

References

- Anderson, R. M. and May, R. M. (1991) *Infectious Disease of Humans: Dynamics and Control*. Oxford: Oxford University Press.
- Andreasen, V., Lin, J. and Levin, S. A. (1997) The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.*, **35**, 825–842.
- Boni, M. J., Gog, J. R., Andreasen, V. and Christiansen, F. B. (2004) Influenza drift and epidemic size: the race between generating and escaping immunity. *Theoret. Popul. Biol.*, **65**, 179–191.
- Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997) Generalized partially linear single-index models. *J. Am. Statist. Ass.*, **92**, 477–489.
- Couch, R. B. and Kasel, J. A. (1983) Immunity to influenza in man. *A. Rev. Microbiol.*, **37**, 529–549.
- Ellner, S. P., Bailey, B. A., Bobashev, G. V., Gallant, A. R., Grenfell, B. T. and Nychka, D. W. (1998) Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *Am. Nat.*, **151**, 425–440.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.

- Ferguson, N. M., Galvani, A. P. and Bush, R. M. (2003) Ecological and immunological determinants of influenza evolution. *Nature*, **422**, 428–433.
- Finkenstädt, B. F. and Grenfell, B. T. (2000) Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Statist.*, **49**, 187–205.
- Fletcher, R. (1987) *Practical Methods of Optimization*, 2nd edn. New York: Wiley.
- Glass, K., Xia, Y. and Grenfell, B. T. (2003) Interpreting time-series analysis for continuous-time biological models—measles as a case study. *J. Theoret. Biol.*, **223**, 19–25.
- Gog, J. R. and Grenfell, B. T. (2002) Dynamics and selection of many-strain pathogens. *Proc. Natn. Acad. Sci. USA*, **99**, 17209–17214.
- Grenfell, B. T., Bjørnstad, O. N. and Kappey, J. (2001) Travelling waves and spatial hierarchies in measles epidemics. *Nature*, **414**, 716–723.
- Grenfell, B. T. and Dobson, A. (1995) *Ecology of Infectious Diseases in Natural Populations*. Cambridge: Cambridge University Press.
- Grenfell, B. T. and Harwood, J. (1997) (Meta)population dynamics of infectious diseases. *Trends Ecol. Evoln*, **12**, 395–399.
- Gupta, S., Swinton, J. and Anderson, R. M. (1994) Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. Lond. B*, **256**, 231–238.
- Hethcote, H. W. and Yorke, J. A. (1984) Gonorrhea transmission dynamics and control. *Lect. Notes Biomath.*, **56**.
- Hochberg, M. E. (1991) Nonlinear transmission rates and the dynamics of infectious diseases. *J. Theoret. Biol.*, **153**, 301–321.
- Liew, C. K. (1976) Inequality constrained least-squares estimation. *J. Am. Statist. Ass.*, **71**, 746–751.
- Liu, R. M., Hethcote, H. W. and Levin, S. A. (1987) Dynamical behaviour of epidemiological models with nonlinear incidence rates. *J. Math. Biol.*, **25**, 359–380.
- Murphy, B. R. and Clements, M. L. (1989) The systemic and mucosal immune response of humans to influenza A virus. *Curr. Top. Microbiol. Immunol.*, **146**, 107–116.
- Nicholson, K. G., Webster, R. G. and Hay, A. J. (1998) *Textbook of Influenza*. Oxford: Blackwell Science.
- Pease, C. M. (1987) An evolutionary epidemiological mechanism, with applications to type A influenza. *Theoret. Popln Biol.*, **31**, 422–452.
- Tong, H. (1990) *Nonlinear Time Series Analysis: a Dynamical System Approach*. London: Oxford University Press.
- Xia, Y. and Härdle, W. (2002) Semi-parametric estimation of the generalized partially single-index linear models. *Discussion Paper 56-2002*. Sonderforschungsbereich 373, Humboldt University, Berlin.