

# Estimation of Infectious Disease Parameters from Serological Survey Data: The Impact of Regular Epidemics

Whitaker H.J.<sup>1</sup> and Farrington C.P.<sup>1</sup>

January 26, 2004

<sup>1</sup> Department of Statistics, The Open University, Milton Keynes, UK.

Corresponding author:

C.P. Farrington,

Department of Statistics,

Faculty of Mathematics and Computing,

The Open University,

Walton Hall,

Milton Keynes,

MK7 6AA

UK

Tel: +44 (0) 1908 654840

E-mail: [c.p.farrington@open.ac.uk](mailto:c.p.farrington@open.ac.uk)

Source of Funding: The Wellcome Trust, project grant 061830

## Summary

Serological surveys are a useful source of information about epidemiological parameters for infectious diseases. In particular they may be used to estimate contact rates, forces of infection, the reproduction number, and the critical vaccination threshold. However, these estimation methods require the assumption that the infection is in endemic equilibrium. Such equilibria seldom exist in practice: for example, many common infections of childhood exhibit regular epidemic cycles. In this paper we investigate whether ignoring such cycles produces biased estimates. We apply the methods to data on mumps and rubella in the U.K. prior to the introduction of the combined measles, mumps, rubella (MMR) vaccine. We conclude that past epidemics have only a marginal effect on estimates, and that standard methods that do not adjust for regular epidemics are valid.

Key words: critical vaccination threshold, epidemic, force of infection, infectious disease, reproduction number, sensitivity analysis, serological survey.

# 1 Introduction

Epidemic models are commonly used for understanding the epidemiology of infections and guiding policy decisions on vaccination programmes. Contrasting methods have been used for estimating parameters of interest, such as contact rates, the force of infection, and derived quantities such as the reproduction number and the critical vaccination threshold. For example, estimation of reproduction numbers may be based on serological survey data [1], time series data [2] or outbreak data [3]; for a review see [4]. However, extensive and reliable times series or outbreak data are lacking for many infections. In recent years, most of the models used in a policy-making context have been based on serological survey data, which are now widely available for many infections [5, 6]. The widespread use of such models motivates the present paper and provides our focus.

The survival methods commonly used with serological survey data assume that the infection is in endemic equilibrium [1, 7, 8, 9]. This implies that the infection hazard is stationary over time. In fact, endemic equilibria are convenient limiting devices that seldom apply in reality. For example, in the absence of vaccination, many childhood infections undergo regular epidemic cycles. This is the case for measles, mumps and rubella, which had epidemic periods of 2 to 5 years in the U.K. prior to mass vaccination [1]. The timing of these epidemic cycles affect an individual's immunity. For example, children aged 2 years are less likely to have been infected with rubella if they were born just after an epidemic rather than just before. Thus, the question arises as to how much the presence of epidemic fluctuations affects the validity of estimates derived from serological survey data under equilibrium assumptions. For example, it has been suggested that ignoring temporal variation in the numbers of susceptibles would underestimate

reproduction numbers [10].

In this paper we undertake a sensitivity analysis of the endemic equilibrium assumption. We compare parameter estimates obtained under standard (but clearly invalid) equilibrium assumptions, with those obtained from extended models which take into account regular epidemic cycles. As is well known, however, age and time effects are to a large extent confounded in a single serological survey [11, 12]. In order to model age and epidemic effects jointly, we augment the serological data with case report data to provide independent information on the timing of epidemics. Combining cross-sectional and time series data in this way has also been used in other fields [13].

Joint modelling of age and time effects in the context of infectious disease modelling has been discussed by several authors [12, 14, 15, 16, 17]. Our focus, however, is different: temporal effects are not our primary interest. Rather, we are interested in the sensitivity of our models to such effects, which we regard as nuisance parameters. Thus, we use a descriptive modelling approach to account for epidemic fluctuations. This contrasts with the more analytical approach used in studies of the dynamics of infection, for example the time series susceptible infected recovered (TSIR) model [2, 18, 19], or models for seasonal effects in transmission [20, 21, 22], which seek to model the underlying transmission process over time.

We focus on modelling mumps and rubella infections prior to the introduction of the combined measles, mumps, rubella (MMR) vaccine in the UK. These data are described in section 2. The epidemic models allowing for epidemic cycles are introduced in section 3. In section 4 we describe estimation methods. In section 5 we apply these models to the rubella and mumps data, and compare estimates of the average force of infection, the

basic reproduction number  $R_0$  and the critical vaccination threshold  $V_c$  obtained with and without allowance for epidemic cycles. Finally in section 6 we discuss the models and the results.

## 2 Data

Age-stratified seroprevalence data for mumps and rubella in the UK were collected between November 1986 and December 1987, to provide baseline information prior to the introduction of the MMR vaccine [23]. The data were previously analysed to estimate forces of infection [24] and estimate basic reproduction numbers [9]. These analyses rest on the assumption that mumps and rubella were in endemic equilibrium in the period leading up to the survey. The serological survey data used in the paper have been published previously [9].

The rubella data include only boys since girls routinely received rubella vaccine at age 10-15 years at that time [1]. Thus the epidemiological parameters we estimate for rubella are derived from a partially vaccinated population. Consequently, the force of infection and reproduction number will be less than if the population were vaccine-free. Mathematical modelling suggests that this effect is small ([1], p.106-8). Our focus, in any case, is to investigate the sensitivity of the results to the endemic equilibrium assumption, rather than obtain estimates valid for a vaccine-free population.

Epidemic cycles are reflected in serological profiles, especially at younger ages. Thus, in theory, it is possible to correct for epidemicity directly using only serological data. However, in practice such adjustments are unreliable because they are sensitive to small fluctuations in the younger age groups. Therefore, we augmented the serological data with data on case reports. Such data are available from a variety of sources, for example statutory

notifications and sentinel reporting schemes. We used laboratory reports for mumps and rubella from England and Wales, 1975-1987, obtained from the Communicable Disease Surveillance Centre.

These data are given in Table 1. Laboratory reports are obtained from diagnostic tests, and therefore are not representative of the age or other structure of the population, and are subject to variations in reporting efficiency. However, they reliably indicate the temporal pattern of epidemics. The laboratory data show that epidemics occurred with periods of about 3 years for mumps and 4 years for rubella; the last epidemic peaks prior to the serological survey occurred in 1984 for mumps and 1983 for rubella.

### 3 Models

We consider endemic immunising infections in a large population, directly spread from person to person, with short latency and infectious periods. Such infections generally exhibit epidemic cycles [1]. For simplicity, we assume that mortality in the population is of type I, i.e. all individuals live to age  $L$ , the life expectancy, and then die; our methods may be generalised to arbitrary mortality functions. Transmission of the infection depends on the age-specific contact function  $\beta(x, y)$ , which is the average per capita rate at which an individual of age  $y$  makes contact with individuals of age  $x$ . We assume that the contact function is stationary over time; this is reasonable on the annual time scale of the data. On shorter time scales the contact function may vary, for example seasonally, but such variation is not considered here.

### 3.1 Force of infection

Let  $I(y, t)$  denote the number of infectives of age  $y$  at time  $t$ , and let  $\lambda(x, t)$  denote the hazard of infection (or force of infection), i.e. the rate at which susceptible individuals of age  $x$  acquire infection at time  $t$ . These quantities are related by the integral equation

$$\lambda(x, t) = \int_0^L \beta(x, y) I(y, t) dy.$$

In line with both theory and observation, the number of infectives varies cyclically over time. Typically, there are slight differences in phase between age groups, related to differences in contact rates. These phase differences are of the order of  $D + D'$ , where  $D$  is the infectious period and  $D'$  the latent period. For infections with  $D + D'$  much less than one year, phase differences are ignorable. We therefore model  $I(y, t)$  by independent age and temporal components:

$$I(y, t) \simeq I_0(y)\omega(t)$$

where  $\omega(t)$  is periodic with period  $T$ , that is,  $\omega(t + T) = \omega(t)$  for all  $t$ . The force of infection  $\lambda(x, t)$  can then be written

$$\begin{aligned} \lambda(x, t) &\simeq \left[ \int_0^L \beta(x, y) I_0(y) dy \right] \omega(t) \\ &= \lambda_0(x) \omega(t). \end{aligned}$$

To identify the components, we require that

$$\frac{1}{T} \int_0^T \omega(u) du = 1.$$

Thus,

$$\lambda_0(x) = \int_0^L \beta(x, y) I_0(y) dy \tag{1}$$

is the average age at infection, averaged over epidemic cycles. The probability that an individual of age  $a$  is susceptible at time  $t$  is

$$S(a, t) = \exp \left\{ - \int_0^a \lambda(u, t - a + u) du \right\} \tag{2}$$

$$= \exp \left\{ - \int_0^a \lambda_0(u) \omega(t - a + u) du \right\}.$$

Note that the quantity  $\exp \left\{ - \int_0^a \lambda_0(u) du \right\}$  is the geometric mean of  $S(a, t)$  over an epidemic period  $T$ , since

$$\exp \left\{ - \int_0^a \lambda_0(u) du \right\} = \exp \left\{ \int_0^T \log S(a, t) dt \right\}^{1/T} \quad (3)$$

Thus averaging the force of infection corresponds to taking geometric means of the proportions susceptible. Theoretical considerations indicate that this, rather than arithmetic averaging of the proportions susceptible, is required to avoid bias [10]. Our focus is therefore on estimating the average force of infection  $\lambda_0(x)$ .

### 3.2 Epidemiological parameters

We wish to estimate  $\lambda_0(x)$  and  $\omega(t)$ , but also derived quantities including the basic reproduction number  $R_0$  and the critical vaccination threshold  $V_c$  [4, 9, 25].

The basic reproduction number  $R_0$  is the average number of individuals infected by one infective introduced into a host population where everyone is susceptible. If  $R_0 > 1$  the infection may become endemic, whereas if  $R_0 \leq 1$  the infection will eventually die out.  $R_0$  is a key parameter in infectious disease epidemiology: the higher the value of  $R_0$ , the harder it will be to control the infection.

$R_0$  is the leading eigenvalue of  $\beta^*(x, y)$  where

$$\beta^*(x, y) = \frac{ND}{L} \beta(x, y).$$

Here  $N$  is the population size,  $D$  the average infectious period,  $L$  the life expectancy.

Suppose now that a long-term vaccination programme is implemented in the population. Let  $\sigma_V(x)$  denote the proportion of individuals of age  $x$  who

remain unprotected by vaccination. Thus  $\sigma_V(x)$  depends on the proportion vaccinated, the vaccination schedule and the efficacy of the vaccine. Let  $R_V$  denote the leading eigenvalue of  $\sigma_V(x)\beta^*(x, y)$ . If  $R_V \leq 1$ , endemic transmission cannot be sustained and the infection is said to have been eliminated from the population. On the other hand, if  $R_V > 1$  the infection remains endemic. The critical vaccination threshold,  $V_c$ , is the minimum proportion of the population that must be vaccinated under the programme in order to reduce  $R_V$  to 1.

## 4 Estimation from serological data and case reports

In this section we describe how the quantities introduced in Section 3 may be estimated, taking into account epidemic cycles. The overall likelihood consists of two components: the likelihood for the case reports data, and the likelihood for the serological data. The case reports data provide information on  $\omega(t)$ , whereas the serological data provide information on both  $\lambda_0(x)$  and  $\omega(t)$ .

### 4.1 Likelihood for case reports

We begin with the case reports data. The expected number of cases reported at time  $t$ ,  $\mu(t)$ , depends on the number of infectives in each age group,  $I(x, t)$ , and the reporting efficiency, which we assume to be of the form  $\kappa(x, t)$ . Thus

$$\begin{aligned}\mu(t) &= \int_0^L \kappa(x, t)I(x, t)dx \\ &\simeq \omega(t) \int_0^L \kappa(x, t)I_0(x)dx \\ &= \omega(t)\kappa(t)\end{aligned}$$

where  $\kappa(t) = \int_0^L \kappa(x, t) I_0(x) dx$ . In order to identify the epidemic cycles we must assume that only  $\omega(t)$  is periodic. Thus we assume that  $\kappa(t)$  represents a secular trend, for example due to changes in diagnostic techniques or in the coverage of the surveillance system. In our application we chose  $\kappa(t) = \rho e^{-\xi t}$ . We modelled the periodic variation  $\omega(t)$  by

$$\omega(t) = \frac{2\pi}{\phi} \frac{\exp\{\tau \cos(\phi t + \psi)\}}{\int_0^{\frac{2\pi}{\phi}} \exp\{\tau \cos(\phi u)\} du}. \quad (4)$$

The parameter  $\tau$  varies with the amplitude of the epidemics, which have period  $T = 2\pi/\phi$ . The phase parameter  $\psi$  determines the timing of the serological survey during the epidemic cycle.

We assumed the number of cases reported at time  $t$ ,  $Z_t$ , was distributed negative binomial with shape  $\nu$  and mean  $\mu(t)$  and hence variance  $\mu(t) + \mu(t)^2/\nu$ . Thus, the log likelihood kernel  $l_C$  for the case reports is:

$$\begin{aligned} l_C &= \sum_t \log \Gamma(Z_t + \nu) - \log \Gamma(\nu) + \nu \log \nu \\ &\quad + Z_t \log \mu(t) - (Z_t + \nu) \log(\mu(t) + \nu) \end{aligned}$$

and involves the six parameters  $\rho, \xi, \nu, \tau, \phi, \psi$ .

## 4.2 Likelihood for serological data

The serological survey data provide information on both  $\lambda_0(x)$  and  $\omega(t)$ . We used two models for the average age-specific force of infection  $\lambda_0(x)$ , namely the gamma function

$$\lambda_0(x) = \alpha x^\delta \exp(-x/\gamma) \quad (5)$$

and the piecewise constant model  $\lambda_0(x) = \lambda_{0i}$ ,  $x \in (a_{i-1}, a_i]$ ,  $i = 1, 2, \dots, 5$  with  $a_0 = 0, a_5 = L$ . Suppose that a serological survey is undertaken at time  $t_s$ . If  $r_a$  out of  $n_a$  individuals of age  $a$  were seropositive, we modelled

$$r_a \sim \text{Binomial}(n_a, 1 - S(a, t_s))$$

where  $S(a, t_s)$  is given in equation (2). The log likelihood kernel  $l_S$  for the serological survey data is therefore

$$l_S = \sum_a \{r_a \log(1 - S(a, t_s)) + (n_a - r_a) \log S(a, t_s)\}.$$

With the gamma force of infection,  $l_S$  involves the six parameters  $\alpha, \delta, \gamma, \tau, \phi, \psi$ . The combined log likelihood kernel for the case reports and serological data is then  $l_C + l_S$  and involves nine parameters (eleven for the piecewise force of infection model).

### 4.3 Estimation of $R_0$

This section contains only a brief description of the method used for estimating the basic reproduction number  $R_0$  [9]. The aim is to estimate  $R_0$  using the average force of infection  $\lambda_0(x)$  derived after allowing for past epidemics, and to compare this with estimates obtained without allowing for past epidemics.

To estimate  $R_0$  we first need a model for the contact function  $\beta(x, y)$ . We used the same four models for  $\beta(x, y)$  as Farrington et al. 2001 [9]. These include the proportional mixing model in which  $\beta(x, y)$  is of the form  $u(x)u(y)$ . For this model,

$$R_0 \simeq \frac{\int_0^L \lambda_0(x)^2 dx}{\int_0^L \lambda_0(x)^2 S_0(x) dx}$$

where  $\lambda_0(x)$  is the average force of infection, modelled using a gamma function, and  $S_0(x) = \exp\{-\int_0^x \lambda_0(u) du\}$ . As shown above, this corresponds to the geometric mean of  $S(x, t)$  over an epidemic period.

The other three models are matrix models in which the contact rates are assumed constant within the 5 age groups  $(a_{i-1}, a_i], i = 1, \dots, 5$ . Thus, for  $x \in (a_{i-1}, a_i], y \in (a_{j-1}, a_j]$ , we assume that  $\beta(x, y)$  is constant and define  $\beta_{ij} = (ND/L)\beta(x, y)$ . Let  $B$  denote the matrix with entry  $\beta_{ij}$  in

position  $(i, j)$ . Then  $R_0$  is the leading eigenvalue of  $MB$  where  $M$  is the diagonal matrix of age group widths. The matrix  $B$  is estimated by solving the equilibrium equation based on the average force of infection:

$$\lambda_0(x) \simeq \frac{ND}{L} \int_0^L \beta(x, y) \lambda_0(y) S_0(y) dy. \quad (6)$$

When  $\beta(x, y)$  is a  $5 \times 5$  matrix, the average force of infection  $\lambda_0(x)$  is piecewise constant on each of the 5 age groups. To resolve the unidentifiability inherent in (6), only 5 free parameters are allowed. We use the models A, B, C shown in Figure 1.

#### 4.4 Estimation of $V_c$

The same four contact functions were used to estimate the critical vaccination threshold, under the MMR vaccination schedule currently in use in the U.K. Under this schedule, one dose is given at 15 months and one at 4 years of age. We assumed that exactly the same children receive the second dose as received the first [26]. Denote by  $v$  the proportion of the population who receive the vaccine. After receiving the first dose of vaccine, some proportion  $w$  of vaccinated children become immune. We assumed that, on administration of the second dose at age 4, a proportion  $w$  of those who remained unprotected after receiving a first dose acquire immunity. The probability  $\sigma_V(x)$  that an individual of age  $x$  is unprotected by the MMR vaccination programme is then

$$\sigma_V(x) = \begin{cases} 1 & 0 \leq x < 1.25 \\ 1 - wv & 1.25 \leq x < 4 \\ 1 - wv(2 - w) & 4 \leq x < 75. \end{cases}$$

For both mumps and rubella we took  $w = 0.95$ . For the proportional mixing model we have

$$R_v \simeq \frac{\int_0^L \lambda_0(x)^2 \sigma_V(x) dx}{\int_0^L \lambda_0(x)^2 S_0(x) dx}$$

Setting  $R_v = 1$ , and  $w = 0.95$  in  $\sigma_V(x)$  and rearranging, we obtain the following explicit expression for the critical vaccination threshold:

$$V_c \simeq \frac{\int_0^L \lambda_0(x)^2 dx - \int_0^L \lambda_0(x)^2 S_0(x) dx}{0.95 \int_{1.25}^4 \lambda_0(x)^2 dx + 0.9975 \int_4^L \lambda_0(x)^2 dx}.$$

For each matrix models  $B$  we first calculated  $M_i(v)$ , the average number of years spent unprotected in age group  $i$ ,  $i = 1, 2..5$ , and let  $M(v) = \text{Diag}(M_1(v), \dots, M_5(v))$ . Then  $R_v$  is the leading eigenvalue of  $M(v)B$ , and the critical vaccination threshold  $V_c$  is the value of  $v$  for which  $R_v = 1$ .

#### 4.5 Goodness of fit, standard errors and confidence intervals

The goodness of fit of the models was assessed by the Pearson  $\chi^2$  statistic [27]. For the case reports data this is

$$\sum_t \frac{(Z_t - \mu(t))^2}{\mu(t) + \nu^{-1} \mu(t)^2}.$$

For the serological survey data, the Pearson  $\chi^2$  is

$$\sum_a \frac{\{r_a - n_a(1 - S(a, t_s))\}^2}{n_a S(a, t_s)(1 - S(a, t_s))}.$$

The overall  $\chi^2$  is obtained by summing over data sets.

Also, a likelihood ratio test of  $\xi = 0$  was carried out to test whether the term  $e^{-\xi t}$ , which describes a secular trend in the case reports data, should be dropped from the model.

Approximate standard errors for the parameters and 95% percentile confidence intervals for  $R_0$  and  $V_c$  were obtained by non-parametric bootstrapping. For the serological survey data we resampled cases within age groups; for the case reports data we resampled Pearson residuals [28]. For each model we drew 599 samples.

## 5 Application to mumps and rubella data

The time  $t_s = 0$  was set at 31 December 1987, at the end of the sampling period for the serological survey, and we took  $L = 75$  years. The age group cut points were 3, 8, 15, 25 years.

There was little evidence of a secular trend in the rubella case reports data ( $\chi^2 = 1.7$ ,  $p = 0.2$ ), but evidence of a declining trend in the mumps data ( $\chi^2 = 14.6$ ,  $p = 0.0001$ ). The Pearson  $\chi^2$  goodness of fit statistics are shown in Table 2. The models without the epidemic effects were only fitted to the serological survey data, not the case reports data. The contributions to the overall Pearson  $\chi^2$  statistics from the serological data using the models with epidemic effects are also included in Table 2. In general, the fit to the serological data improved for the youngest ages when epidemic cycles were allowed for.

Figure 2 shows the fit to the mumps and rubella laboratory reports using both the gamma function and piecewise constant models for the force of infection. The estimated epidemic periods were 4.33 years (SE 0.15) for rubella and 2.81 years (SE 0.03) for mumps; these estimates were obtained modelling the case reports data alone. There are a few poorly fitting data points, such as the report numbers for 1978, probably due to differences in reporting efficiency from year to year. Figure 3 shows the fit of the gamma function models to the serological survey data.

Table 3 shows force of infection parameter estimates and standard errors, obtained both taking account of, and without taking account of epidemic cycles. The estimated parameters for the piecewise constant models obtained with and without allowing for epidemic cycles are similar, typically lying within one, and always within two, standard errors of each other. Figure 4 shows the estimated average forces of infection for the gamma function

models. Clearly, allowing for epidemic cycles makes hardly any difference to the shape of the average force of infection for mumps, and only a marginal difference for rubella (of 12% at most).

The estimates of  $R_0$  and  $V_c$  for mumps and rubella for the four contact functions (proportional mixing and 3 matrix models), calculated both with and without taking into account epidemic effects, are shown in Tables 4 and 5. With the exception of the proportional mixing model for rubella, the estimates obtained by one method lie within the 95% confidence interval obtained by the other method, and differ by less than 5%.

The models were also fitted allowing for individual frailties using the methods in Farrington et al. (2001) [9]. Again, results showed that allowing for epidemic cycles had little impact on parameter estimates (not shown).

As expected, the serological data contain very little information about the epidemic cycles. Nevertheless, we also attempted to estimate the parameters using solely the serological data by maximising the log-likelihood  $l_S$ . The results for the rubella data were similar to those obtained using the combined data. However, for the mumps data, the model was unstable.

## 6 Discussion

This study was motivated by the concern that commonly used methods for deriving key epidemiological quantities for infectious diseases using serological survey data make no allowance for epidemicity, and rely on a manifestly incorrect stationarity assumption. Under this assumption, the infection should have reached an equilibrium, in which the age-specific force of infection is stationary over time. In reality, many common immunising infections of childhood undergo regular epidemic cycles. Hence the concern that estimation methods based on serological survey data, in which it is

difficult to distinguish age and periodic temporal effects, may be flawed.

To our knowledge, this is the first statistical study of the impact of past epidemics on the validity of analyses of serological survey data under equilibrium assumptions. Our results suggest that their impact is marginal and ignorable. The estimated average forces of infection, and derived quantities such as the reproduction number and critical vaccination threshold, are largely insensitive to epidemic cycles. This is especially true of piecewise constant and matrix models, which are more robust to slight biases than more strongly parametric models such as the proportional mixing model with gamma force of infection.

Interestingly, ignoring epidemic cycles did not appear to underestimate  $R_0$  as predicted [10]. This may be due to slight overestimation of the force of infection in younger age groups (Table 3 and Figure 4).

It might be objected that these results relate only to two data sets, and that substantially different results might have been obtained if the epidemics had occurred at different times. We consider this unlikely: the epidemic periods of mumps, 2.81 years, 95% CI (2.76, 2.87) and rubella, 4.33 years (4.06, 4.62), are quite different, but both data sets yield similar conclusions. We also varied the survey time  $t_s$  to coincide with other points during the data collection period from November 1986 to December 1987. This made little difference to the results.

A heuristic argument suggests that our findings for mumps and rubella apply generally to immunising infections with short latent and infectious periods, i.e. to infections that display epidemic fluctuations. The force of infection is often monotone for the first years of life, to an age of the order of  $A$ , the average age at infection. On the other hand, regular epidemic cycles occur with periods of the order of  $T \simeq 2\pi\sqrt{A(D+D')}$  where  $D$

is the infectious period and  $D'$  the latent period [1]. Typically,  $T$  is not substantially larger than  $A$ , and is often less than  $A$ . Hence, estimates of the force of infection that ignore regular epidemic cycles will tend to represent an ‘average’ value that differs little from the true average force of infection from equation (1). We conclude that standard methods that do not adjust for epidemic cycles are likely to produce valid results in this setting.

We considered only the effect of regular cycles. We suspect that broadly similar conclusions would apply to infections with irregular but frequent epidemics, with average inter-epidemic period less than  $A$ : for such infections, we would expect a similar averaging to occur when estimating the force of infection. However, infections with infrequent epidemics (whether regular or not) may generate biased results. For such infections, it may be necessary to model temporal fluctuations explicitly as we have done here, either using a time series of case reports or multiple serological surveys.

Aggregation of data over spatial units can mask or distort epidemic cycles for some infections. For example, whooping cough has been shown to lack spatial coherence prior to the introduction of mass vaccination in the UK, a phenomenon which masked its 2 to 2.5 year epidemic cycle [29]. Aggregation of serological data corresponds to taking arithmetic rather than geometric means [10]; see equation (3). Unless the component data sets are in phase with regard to epidemic cycles, aggregation may result in the average force of infection being underestimated in younger children. In older age groups, namely those having experienced more than one epidemic cycle, variation in seroprevalence due to past epidemics is less pronounced.

In this paper we considered regular epidemic cycles, and hence assumed that the contact function  $\beta(x, y)$  remains constant from one year to the next. The endemic equilibrium assumption may be violated in other ways:

for example, the birth rate or mortality function, and hence  $N$  or  $L$ , could vary [30]. Alternatively, contact rates may vary over time, reflecting changes in the social or natural environment. Thus  $\beta(x, y)$  could be time-dependent. In such settings  $R_0$  and  $V_c$  would also vary over time. The estimation of such temporal effects over longer time scales is the subject of ongoing research [31].

## **Acknowledgements**

We thank Nigel Gay of the Communicable Surveillance Centre, London, for providing the laboratory data and for useful discussions. The paper was much improved following comments by two referees. This project was supported by project grant 061830 from the Wellcome Trust.

## References

- [1] Anderson RM and May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1991.
- [2] Bjornstad ON, Finkenstadt BF and Grenfell BT. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs* 2002; **72**:169-184.
- [3] Becker NG. *Analysis of Infectious Disease Data*. Chapman and Hall, London, 1989.
- [4] Dietz K. The Estimation of the Basic Reproduction Number for Infectious Diseases. *Statistical Methods in Medical Research* 1993; **2**:23-41.
- [5] Edmunds W, Gay N, Kretzschmar M, Pebody R and Wachmann H. The pre-vaccination epidemiology of measles, mumps and rubella in Europe. *Epidemiology and Infection* 2000; **125**:635-650.
- [6] Edmunds W, Van de Heijden O, Eerola M and Gay N. Modelling rubella in Europe. *Epidemiology and Infection* 2000; **125**:617-634.
- [7] Dietz K. Transmission and control of arboviruses. In: *Epidemiology*, Ludwig D and Cooke KL editors. SIAM, Philadelphia, 1975, pp104-121.
- [8] Grenfell BT and Anderson RM. The Estimation of Age-Related Rates of Infection from Case Notifications and Serological Data *Journal of Hygiene* 1985; **95**:419-436.
- [9] Farrington CP, Kanaan MN and Gay NJ. Estimation of the Basic Reproduction Number for Infectious Diseases from Age-Stratified Serological Survey Data. *Applied Statistics* 2001; **50**:1-33.

- [10] Keeling MJ and Grenfell BT. Individual-based perspectives on  $R_0$ . *Journal of Theoretical Biology* 2000; **203**:51-61.
- [11] Schenzle, D., Dietz, K. and Frosner, G. (1979). Antibody against hepatitis A in seven European countries. *American Journal of Epidemiology* 1979; **110**:70-76.
- [12] Ades A and Nokes D. Modeling age- and time-specific incidence from seroprevalence: toxoplasmosis. *American Journal of Epidemiology* 1993; **137**:1022-1034.
- [13] Besbeas P, Freeman S, Morgan B and Catchpole E. Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* 2002; **58**:540-547.
- [14] Marschner I. Fitting a multiplicative incidence model to age- and time-specific prevalence data. *Biometrics* 1996; **52**:492-499.
- [15] Marschner I. A method for assessing age-time disease incidence using serial prevalence data. *Biometrics* 1997; **53**:1384-1398.
- [16] Nagelkerke N, Heisterkamp S, Borgdorff M, Broekmans J and Van Houwelingen H. Semi-parametric estimation of age-time specific infection incidence from serial prevalence data. *Statistics in Medicine* 1999; **18**:307-320.
- [17] Ferguson N, Donnelly C and Anderson R. Transmission dynamics and epidemiology of dengue: insights from age-stratified sero-prevalence surveys. *Phil. Trans. R. Soc. Lond. B* 1999; **354**:757-768.
- [18] Finkenstadt B. and Grenfell B. Time series modelling of childhood diseases: a dynamical systems approach. *Applied Statistics* 2000; **49**:187-205.

- [19] Finkenstadt B, Bjornstad O and Grenfell B. A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreak. *Biostatistics* 2002; **3**:493-510.
- [20] London WP and Yorke JA. Recurrent outbreaks of measles, chickenpox and mumps I: Seasonal variation in contact rates. *American Journal of Epidemiology* 1973; **98**:453-468.
- [21] Fine PEM and Clarkson JA. Measles in England and Wales I: An analysis of the factors underlying seasonal patterns. *International Journal of Epidemiology* 1982; **11**:5-14.
- [22] Deguen S, Thomas G and Chau N. Estimation of the contact rate in the seasonal SEIR model: application to chickenpox incidence in France. *Statistics in Medicine* 2000; **19**:1207-1216.
- [23] Morgan-Capner P, Wright J, Miller CL and Miller E. Surveillance of Antibody to Measles, Mumps, and Rubella by Age. *British Medical Journal* 1988; **297**:770-772.
- [24] Farrington CP. Modelling Forces of Infection for Measles, Mumps and Rubella. *Statistics in Medicine* 1990; **9**:953-967.
- [25] Diekmann O, Heesterbeek JAP and Metz JAJ. On the Definition and the Computation of the Basic Reproduction Ratio  $R_0$  in Models for Infectious Diseases in Heterogeneous Populations. *Journal of Mathematical Biology* 1990; **28**:365-382
- [26] Wallinga J, Levy-Bruhl D, Gay NJ and Wachmann CH. Estimation of Measles Reproduction Ratios and Prospects for Elimination of Measles by Vaccination in some Western European Countries. *Epidemiology and Infection* 2001; **127**:281-295

- [27] McCullagh P and Nelder JA. *Generalized Linear Models*. Second Edition. Chapman & Hall. London, 1989.
- [28] Davison AC and Hinkley DV. *Bootstrap Methods and their Application*. Cambridge University Press. Cambridge, 1997.
- [29] Rohani P, Earn DJD and Grenfell BT. Opposite patterns of synchrony in sympatric disease metapopulations. *Science* 1999; **286**:968-971.
- [30] Finkenstadt B, Keeling M and Grenfell B. Patterns of density dependence in measles dynamics. *Proceedings of the Royal Society of London, Series B* 1998; **265**:753-762.
- [31] Whitaker HJ and Farrington CP. Modelling the changing epidemiology of childhood infections: application to varicella-zoster. Submitted to *Biometrics* 2003.

## Tables

Table 1: Case report data

year	no. of mumps reports	no. of rubella reports
1975	917	1275
1976	748	650
1977	404	503
1978	1056	3514
1979	571	2399
1980	461	581
1981	729	497
1982	358	1533
1983	436	2533
1984	608	1744
1985	342	805
1986	525	2029
1987	651	1881

Table 2: Pearson  $\chi^2$  goodness of fit statistics ( $\chi^2$  for serological data alone are in brackets)

		Rubella		Mumps	
		epidemic effects			
		Yes	No	Yes	No
Gamma	$\chi^2$	55.2 (41.6)	51.1	64.8 (52.3)	54.1
function	d.f.	49	41	48	41
model	p	0.25	0.13	0.05	0.08
Piecewise	$\chi^2$	55.9 (40.7)	42.1	54.9 (42.4)	46.5
constant	d.f.	47	39	46	39
model	p	0.18	0.34	0.17	0.19

Table 3: Parameter estimates (S.E.) for the force of infection

		Rubella		Mumps	
		epidemic effects			
		Yes	No	Yes	No
Gamma	$\alpha$	0.072 (0.009)	0.090 (0.009)	0.099 (0.009)	0.123 (0.010)
function	$\delta$	0.660 (0.149)	0.547 (0.125)	1.406 (0.142)	1.241 (0.138)
model	$\gamma$	9.040 (1.822)	10.575 (2.222)	3.992 (0.407)	4.283 (0.457)
Piecewise	$\lambda_1$	0.072 (0.010)	0.078 (0.008)	0.084 (0.008)	0.096 (0.007)
constant	$\lambda_2$	0.134 (0.013)	0.135 (0.012)	0.315 (0.014)	0.319 (0.016)
model	$\lambda_3$	0.138 (0.017)	0.138 (0.017)	0.174 (0.022)	0.166 (0.023)
	$\lambda_4$	0.057 (0.016)	0.054 (0.017)	0.052 (0.020)	0.053 (0.019)
	$\lambda_5$	0.033 (0.018)	0.035 (0.018)	0.003 (0.012)	0.003 (0.013)

Table 4: Basic reproduction number and critical immunisation threshold (95% C.I.) for each contact model for rubella

	$R_0$	$R_0$	$V_c$	$V_c$
	epidemic effects			
	yes	no	yes	no
Proportional mixing	2.3 (2.1, 2.5)	2.6 (2.4, 2.9)	0.58 (0.53, 0.62)	0.64 (0.61, 0.67)
Matrix $B_A$	3.3 (2.6, 5.5)	3.3 (2.6, 5.5)	0.72 (0.64, 0.83)	0.73 (0.65, 0.84)
Matrix $B_B$	3.5 (2.8, 5.6)	3.6 (2.8, 5.6)	0.73 (0.67, 0.84)	0.74 (0.67, 0.84)
Matrix $B_C$	4.5 (3.8, 8.9)	4.2 (3.8, 8.7)	0.78 (0.75, 0.89)	0.77 (0.75, 0.89)

Table 5: Basic reproduction number and critical immunisation threshold (95% C.I.) for each contact model for mumps

	$R_0$	$R_0$	$V_c$	$V_c$
	epidemic effects			
	yes	no	yes	no
Proportional mixing	3.1 (3.0, 3.3)	3.2 (3.1, 3.5)	0.69 (0.68, 0.71)	0.71 (0.69, 0.72)
Matrix $B_A$	3.3 (3.1, 3.9)	3.3 (3.1, 4.2)	0.72 (0.70, 0.77)	0.72 (0.71, 0.79)
Matrix $B_B$	8.0 (5.1, 11.3)	8.0 (3.6, 11.4)	0.88 (0.81, 0.91)	0.88 (0.74, 0.91)
Matrix $B_C$	25.5 (4.1, 31.0)	25.5 (4.2, 31.1)	0.96 (0.77, 0.97)	0.96 (0.77, 0.97)

## Figures

Figure 1: Contact matrix structures

$$B_A = \begin{pmatrix} \beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ \beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}$$

$$B_B = \begin{pmatrix} \beta_1 & \beta_1 & \beta_4 & \beta_4 & \beta_5 \\ \beta_1 & \beta_2 & \beta_4 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}$$

$$B_C = \begin{pmatrix} \beta_1 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_2 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_3 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}$$

Figure 2: Case reports data (points) and model fits. Solid line: with gamma force of infection, dashed line: with piecewise constant force of infection.

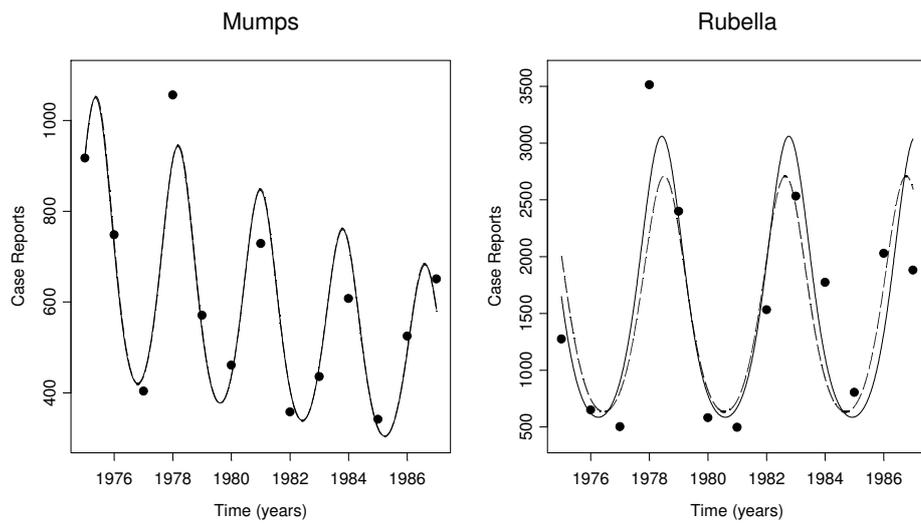


Figure 3: Serological survey data (points) and model fits with gamma force of infection. Solid line: without epidemic effects, dotted line: with epidemic effects.

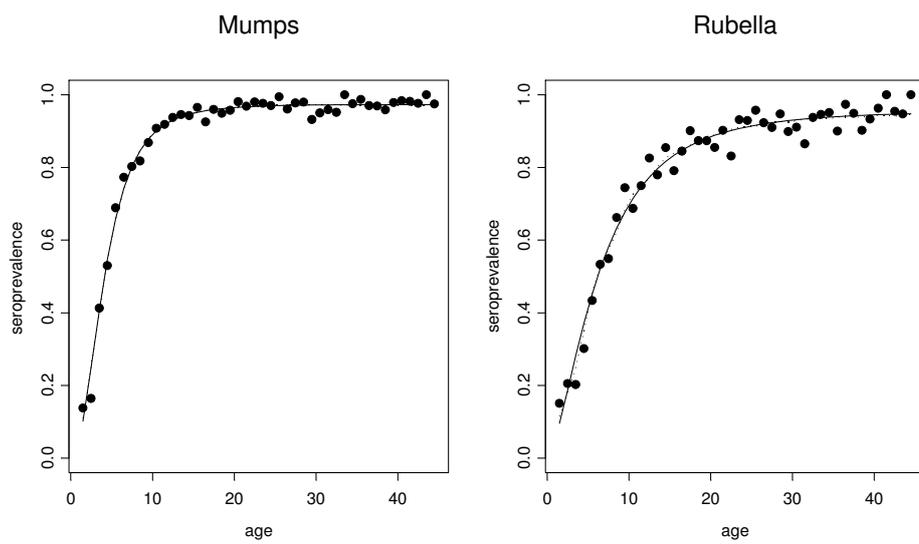


Figure 4: Estimated average force of infection for gamma models. Solid line: without epidemic effects, dotted line: with epidemic effects.

