Biostatistics (2003), **4**, 4, *pp.* 621–632
Printed in Great Britain

Estimation of effective reproduction numbers for infectious diseases using serological survey data

C. P. FARRINGTON*, H. J. WHITAKER
Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
c.p.farrington@open.ac.uk

SUMMARY

The effective reproduction number of an infection, denoted R_e, may be used to monitor the impact of a vaccination programme. If R_e is maintained below 1, then sustained endemic transmission of the infection cannot occur. In this paper we discuss methods for estimating R_e from serological survey data, allowing for age and individual heterogeneity. We describe semi-parametric and parametric models, and obtain an upper bound on R_e when vaccine coverage and efficacy are not known. The methods are illustrated using data on mumps and rubella in England and Wales.

Keywords: Eigenvalue; Frailty; Infectious disease; Reproduction number; Vaccine.

1. INTRODUCTION

Mass vaccination programmes against common childhood infections aim both to protect the individuals vaccinated, and to control the spread of infection in the population. The phenomenon of herd immunity makes it possible to eliminate an infection even though not all individuals are vaccinated. Elimination means reducing the number of susceptibles below a critical threshold, so that spread from an infective cannot produce a large epidemic.

Whether a vaccination programme is achieving elimination at a particular time t after the introduction of mass vaccination is determined by the effective reproduction number of the infection at time t. This quantity, denoted $R_e(t)$, is the average number of infectious individuals resulting from a single infective introduced at time t into the population, given the population mix of vaccine-acquired and naturally acquired immunity at that time. If $R_e(t) \leqslant 1$, then, while infections still occur, for example by limited spread from imported cases, they cannot result in large epidemics. If the value of $R_e(t)$ is greater than 1, or below 1 but on the increase, additional control measures may be called for. Such calculations led to the measles and rubella mass vaccination campaign in 1994 in the UK (Gay *et al.*, 1995). Serological surveillance aims at monitoring the potential for epidemics by estimating $R_e(t)$ at regular time intervals, to assess the overall impact of a vaccination programme.

$R_e(t)$ may be estimated from data on outbreak size or duration, as described by Farrington *et al.* (2003). However, though convenient, these methods are approximate and take no account of heterogeneities in the population. In this paper we consider the estimation of $R_e(t)$ from serological survey data. Existing estimation methods allow for age dependence in contact rates between individuals (Gay *et al.*, 1995; Wallinga *et al.*, 2001). Our aim is to extend these method to incorporate both age-dependent

*To whom correspondence should be addressed.

Biostatistics 4(4) © Oxford University Press; all rights reserved.

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

Table 1. *Paired mumps/rubella antibody test results by age (years), England and Wales, 1996*

Age	−/−	+/−	−/+	+/+	Age	−/−	+/−	−/+	+/+
1	9	0	7	39	23	0	4	2	20
2	4	0	4	35	24	0	2	2	15
3	6	0	6	37	25	0	1	1	11
4	2	1	10	36	26	1	0	2	14
5	0	1	6	42	27	0	2	1	21
6	4	0	4	28	28	2	2	1	16
7	2	1	3	36	29	0	1	0	11
8	2	0	7	40	30	2	3	1	18
9	5	0	11	38	31	0	0	0	11
10	1	0	2	40	32	1	0	1	17
11	2	1	4	32	33	0	2	1	13
12	2	1	8	33	34	1	0	1	16
13	2	1	6	29	35	0	0	0	21
14	0	4	1	38	36	1	2	1	26
15	1	1	2	31	37	0	1	0	15
16	1	3	2	23	38	0	1	1	15
17	1	2	2	24	39	0	0	1	12
18	0	4	4	28	40	0	1	0	9
19	0	5	2	32	41	0	0	0	9
20	0	2	0	16	42	0	1	1	16
21	0	3	0	9	43	0	1	0	11
22	0	1	0	20	44	0	0	0	14

contact rates and random individual heterogeneity: Farrington *et al.* (2001) have shown that individual effects can have a big impact on reproduction numbers.

The paper is organized as follows. In the next section we present the mumps and rubella example we shall use to motivate and illustrate the methods. In Section 3 we review reproduction numbers and define more precisely the effective reproduction number. In Section 4 we describe the methods for estimating $R_e(t)$. In Section 5 we illustrate the methods for the mumps and rubella data, and finally in Section 6 we discuss the methods proposed.

2. MUMPS AND RUBELLA DATA

We illustrate the methods using two sets of serological survey data collected in 1987 and 1996 by the Public Health Laboratory Service (PHLS). In each survey, residual blood samples from males aged 1–44 years of age were tested for antibodies to a variety of childhood infections, including mumps and rubella. For each blood sample, the paired test results for measles and rubella are available. The 1987 data have been published in Farrington *et al.* (2001). The 1996 data, listed by completed year of age, are shown in Table 1.

A positive result for either infection indicates that the individual is immune; a negative result indicates that the individual is susceptible. The 1987 survey was undertaken prior to the introduction of measles, mumps and rubella (MMR) vaccine in October 1988; previously, boys were not vaccinated against mumps or rubella. We shall use the 1987 survey as our first, 'pre-vaccination' sample. The 1996 survey constitutes the second, 'post-vaccination' sample.

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

The MMR vaccination data for each birth cohort from 1987 were obtained from the PHLS website. The recommended age for MMR vaccination is 12–15 months. However, when MMR vaccine was first introduced, children aged 1–4 years were vaccinated. We shall assume that 60% of children born in 1984–86 were vaccinated in the 'catch-up' campaign targeted at 2–4 year-olds. In 1994, a nationwide measles and rubella (MR) campaign was undertaken targeted at children aged 5–16 years. We assume that 90% coverage was achieved in the corresponding birth cohorts (Wiratunga and O'Brien, 1995). We shall take mumps vaccine efficacy to be 85%, and rubella vaccine efficacy 95%. Later in the paper, we shall estimate vaccine coverage in the MMR catch-up campaign and the efficacy of mumps and rubella vaccines.

3. Reproduction numbers

Consider a large population of constant size N, with age-specific mortality rate $\mu(x)$ and age density $m(x)$. We write $M(x) = Nm(x)$. Consider an infection in this population, directly transmitted from person to person, with ignorable infection-related mortality. Transmission depends on contacts between individuals, the nature of which is determined by the route of infection. For example, measles and rubella are transmitted by inhalation of airborne droplets or by direct contact, so for these infections 'close proximity' constitutes a contact. For sexually transmitted infections, contact means sexual contact. Later in the paper we shall exploit the fact that different infections may share similar modes of transmission.

3.1 *Heterogeneity and reproduction numbers*

Let $\beta(x, y)\,\mathrm{d}x$ denote the average per-capita number of effective contacts between an individual of age y and individuals of ages $[x, x + \mathrm{d}x)$. An effective contact is a contact between individuals A and B such that, if A were infective and B susceptible, then A would infect B. We define the cumulative contact rate to be

$$\beta^{+}(x, y) = \int_0^\infty \beta(x, y + s)\left\{1 - F_y(s)\right\} \exp\left\{ - \int_y^{y+s} \mu(t)\,\mathrm{d}t \right\} \mathrm{d}s$$

where $F_y(s)$ is the cumulative distribution function of the infectious period for an individual infected at age y. Thus $\beta^{+}(x, y)\,\mathrm{d}x$ is the total number of effective contacts during the infectious period of an individual who becomes infected at age y, with individuals of age $[x, x + \mathrm{d}x)$. We shall assume that the mean infectious period D is short and independent of the age at infection, so that $\beta^{+}(x, y) \simeq D\beta(x, y)$. If the population were totally susceptible, then the total number of individuals directly infected by a single infectious individual of age y would be

$$\int_0^\infty M(x)\beta^{+}(x, y)\,\mathrm{d}x.$$

A fundamental quantity in infectious disease epidemiology is the basic reproduction number R_0, defined as the expected number of infectious individuals directly infected by a single 'typical' infective in a wholly susceptible population. Diekmann *et al.* (1990) showed that $R_0 = \rho\{M(x)\beta^{+}(x, y)\}$, where $\rho\{A(x, y)\}$ denotes the leading eigenvalue of $A(x, y)$.

Following Farrington *et al.* (2001), we elaborate the model further to include the effect of individual heterogeneity, represented by a continuous, positive random activity level variable U with mean 1. We assume that activity levels are independent of age and act multiplicatively on the contact rate. Thus we expand the contact rate to a function $\beta(x, u; y, v) = uv\beta(x, y)$, where $\beta(x, u; y, v)\,\mathrm{d}x\,\mathrm{d}u$ denotes

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

the average per-capita number of contacts between an individual of age y and activity level v and individuals of ages $[x, x+\mathrm{d}x)$ and activity levels $[u, u+\mathrm{d}u)$. If the infectious period is independent of the activity level, we have $\beta^+(x, u; y, v) \simeq uv\beta^+(x, y)$. Let $f(u)$ denote the density of U. In the presence of heterogeneity, the basic reproduction number is the leading eigenvalue of $M(x)f(u)uv\beta^+(x, y)$ considered as a bivariate function with arguments (x, u) and (y, v). It follows that

$$R_0 = \{1 + \mathrm{var}(U)\} \times \rho\{M(x)\beta^+(x, y)\}.$$

Individual heterogeneity induces a frailty on the pre-vaccination hazard of infection. To see this, let $I^{pre}(y, v)$ denote the number of infectives of age y and activity level v. The superscript pre indicates that the superscripted quantity relates to the period prior to the introduction of mass vaccination. We assume throughout that, prior to the introduction of vaccination, the population is in endemic equilibrium, and hence the average number of infectives in each age group is constant over time; ignoring epidemic fluctuations has a negligible effect on parameter estimates (Whitaker and Farrington, submitted). Thus the pre-vaccination hazard of infection is

$$\begin{aligned}\lambda^{pre}(x, u) &= \int_0^\infty \int_0^\infty uv\beta(x, y)I^{pre}(y, v)\,\mathrm{d}y\,\mathrm{d}v \\ &= u\lambda^{pre}(x).\end{aligned}$$

3.2 *The effective reproduction number R_e*

Suppose now that a vaccination programme is introduced, and some fixed time τ after its introduction a proportion $P(x)$ of individuals of age x are susceptible. ($P(x)$ and other quantities to be defined depend on τ, but we have not made this dependence explicit to avoid cluttering the notation.) The effective reproduction number is the expected number of infectives that would be directly infected by a single 'typical' infective introduced into this population. Thus, in the absence of individual heterogeneity, the effective reproduction number R_e is the leading eigenvalue of $M(x)P(x)\beta^+(x, y)$. The product $M(x)P(x)$ is just the number of susceptible individuals of age x, which may be estimated directly in a serological survey, without needing to know who is or is not vaccinated.

In the presence of individual heterogeneity, however, matters are not so simple. As was the case pre-vaccination, individual heterogeneity induces a frailty on the hazard (or force) of infection. If $I_t^{post}(y, v)$ denotes the number of infectives of age y and activity level v at time t after the introduction of mass vaccination (whence the superscript $post$), then the hazard of infection acting on individuals of age x and activity level u, not protected by vaccination, is

$$\begin{aligned}\lambda^{post}(x, u, t) &= \int_0^\infty \int_0^\infty uv\beta(x, y)I_t^{post}(y, v)\,\mathrm{d}y\,\mathrm{d}v \\ &= u\lambda^{post}(x, t).\end{aligned}$$

In contrast, the hazard of infection acting on individuals protected by vaccination is zero: we assume here that vaccination either protects completely against infection, or imparts no protection whatsoever, and in either case has no effect on infectiousness. Let $\pi(x)$ denote the proportion of individuals of age x who have been vaccinated by time τ and are protected: that is, the proportion with vaccine-induced immunity. We write $M_V(x) = M(x)\{1 - \pi(x)\}$ to denote the number of individuals aged x who are unprotected by

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

vaccination. The proportion of unprotected individuals of age x and activity level u who remain uninfected at time τ after the introduction of mass vaccination is

$$S^{post}(x,u) = \exp\left\{-\int_0^x \lambda^{post}(s,u,\tau-x+s)\,\mathrm{d}s\right\} = \{S^{post}(x)\}^u$$

where $S^{post}(x) = \exp\left\{-\int_0^x \lambda^{post}(s,\tau-x+s)\,\mathrm{d}s\right\}$. Note that the post-vaccination force of infection is time-dependent, as the pre-existing equilibrium will have been destroyed by the introduction of mass vaccination. In the presence of heterogeneity, the effective reproduction number R_e at time τ after the introduction of mass vaccination is the leading eigenvalue of $M_V(x)\{S^{post}(x)\}^u f(u)uv\beta^+(x,y)$. As shown in Farrington (2003), this is equal to $\{1 + \mathrm{var}(U)\}$ times the leading eigenvalue of $M_V(x)\overline{S}^{post}(x)\beta^+(x,y)$, where

$$\overline{S}^{post}(x) = \frac{\int_0^\infty u^2 S^{post}(x)^u f(u)\,\mathrm{d}u}{\int_0^\infty u^2 f(u)\,\mathrm{d}u}. \tag{3.1}$$

Recall that, in the absence of individual heterogeneity, the effective reproduction number R_e is the leading eigenvalue of $M(x)P(x)\beta^+(x,y)$. Here, $P(x)$ is the proportion of individuals of age x who remain susceptible at time τ. Since $P(x) = \{1-\pi(x)\}S^{post}(x)$, R_e is the leading eigenvalue of $M_V(x)S^{post}(x)\beta^+(x,y)$. Thus, unlike R_0, the effect of heterogeneity is not just to inflate the reproduction number by $1 + \mathrm{var}(U)$, but also to replace $S^{post}(x)$ by $\overline{S}^{post}(x)$, rather than by the expectation of $\{S^{post}(x)\}^U$. As a result, since $M_V(x)\overline{S}^{post}(x)$ is not the observed number of susceptibles of age x, a little more work is required in order to estimate R_e.

4. Estimation of R_e from serological survey data

We consider estimation of R_e at some time τ after the introduction of mass vaccination. As before, for clarity we will make no reference to τ in subsequent notation. Two surveys are required: a survey undertaken prior to the introduction of mass vaccination, and a second survey at time τ after its introduction.

In the absence of heterogeneity, the first survey is used to estimate $M(x)\beta^+(x,y)$, and the second to estimate the age-specific proportion susceptible $P(x)$. In the presence of individual heterogeneity, the pre-vaccination survey is also used to estimate $M(x)\beta^+(x,y)$. However, we now need paired data on two infections transmitted by the same route. The method has been described in Farrington *et al.* (2001).

Consider two infections $i = 1, 2$ transmitted by the same route, in an unvaccinated population. To each individual corresponds a random activity level U, which applies to the transmission of both infections since they are transmitted by the same route. We shall assume that the activity levels U are gamma distributed with mean 1 and variance θ^{-1}.

Typically, we assume that contact rates are constant within k age groups, so that the $\beta_i^+(x,y)$ are k-dimensional matrices with only k distinct entries to ensure identifiability. Since the two infections are transmitted by the same route, it is reasonable to assume that $\beta_1^+(x,y)$ and $\beta_2^+(x,y)$ have the same structure. The function $M(x)$ describes the age structure of the population and is assumed known.

Given paired serological data from the 'pre-vaccination' survey, Farrington *et al.* (2001) derive the log likelihood l^{pre} from which estimates of $M(x)\beta_i^+(x,y)$ and θ may be obtained.

In order to estimate the effective reproduction numbers R_{ei} for each of the two infections $i = 1, 2$, we combine the pre-vaccination survey data with data from the second, 'post-vaccination' serological survey. We consider three cases, according to the information available.

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

4.1 *Semi-parametric estimation of R_e*

The second survey is conducted at time τ after mass vaccination was introduced. We assume that, at that time, the proportion of individuals of age x with vaccine-induced protection against infection i, $\pi_i(x)$, is known. This is the case, for example, if vaccine coverage and vaccine efficacy are both known. The probability $P_i(x)$ that an individual of age x remains susceptible to infection i at time τ is

$$P_i(x) = \{1 - \pi_i(x)\} \int_0^\infty \{S_i^{post}(x)\}^u f(u)\, \mathrm{d}u.$$

Thus, with a gamma frailty,

$$P_i(x) = \{1 - \pi_i(x)\} \left\{ 1 + \frac{\Lambda_i^{post}(x)}{\theta} \right\}^{-\theta}$$

where $\Lambda_i^{post}(x) = \int_0^x \lambda^{post}(z, \tau - x + z)\, \mathrm{d}z$ is the baseline cumulative force of infection experienced by an individual of age x in the post-vaccination survey. From equation (3.1) we therefore have

$$\begin{aligned} \overline{S}_i^{post}(x) &= \left\{ 1 + \frac{\Lambda_i^{post}(x)}{\theta} \right\}^{-(\theta+2)} \\ &= \left\{ \frac{P_i(x)}{1 - \pi_i(x)} \right\}^{1+\frac{2}{\theta}} \end{aligned}$$

and

$$\{1 - \pi_i(x)\}\overline{S}_i^{post}(x) = P_i(x) \left\{ \frac{P_i(x)}{1 - \pi_i(x)} \right\}^{\frac{2}{\theta}}.$$

Note that if there is no individual heterogeneity, $\theta \to \infty$ and so $\{1 - \pi_i(x)\}\overline{S}_i^{post}(x) = P_i(x)$. In the presence of heterogeneity, $\theta < \infty$ and the effective reproduction number R_{ei} is $\{1 + \theta^{-1}\}\rho \left\{ M(x) P_i(x) \left\{ \frac{P_i(x)}{1-\pi_i(x)} \right\}^{\frac{2}{\theta}} \beta_i^+(x, y) \right\}$.

Since some individuals are immune as the result of natural infection, $P_i(x) \leqslant 1 - \pi_i(x)$. If, in the post-vaccination survey, n_{0ix} individuals of age x out of n_{ix} tested for antibodies to infection i are susceptible, then we estimate

$$\widehat{P}_i(x) = \min \left\{ \frac{n_{0ix}}{n_{ix}}, 1 - \pi_i(x) \right\}.$$

We then estimate

$$\widehat{R}_{ei} = \{1 + \widehat{\theta}^{-1}\}\rho \left\{ M(x) \widehat{P}_i(x) \left\{ \frac{\widehat{P}_i(x)}{1 - \pi_i(x)} \right\}^{\frac{2}{\theta}} \widehat{\beta}_i^+(x, y) \right\}$$

where $\widehat{\theta}$ and the $M(x)\widehat{\beta}_i^+(x, y)$ are obtained from the 'pre-vaccination' survey. Note that this method does not require paired post-vaccination data on both infections, and hence could be described as semi-parametric: a parametric model is used to estimate $M(x)\beta_i^+(x, y)$ and θ from pre-vaccination serological survey data, but no model is assumed for the post-vaccination data. Parametric estimation of R_{ei} is discussed below.

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

4.2 *Parametric estimation of R_e*

If paired data on the same infections are available in both the first and second surveys, then more explicit modelling of $P_i(x)$ is possible. As before, label the two infections as $i = 1, 2$ and write $\pi_i(x)$ for the proportion of individuals of age x with vaccine-induced immunity to infection i. Also let $\pi_{12}(x)$ denote the proportions of individuals of age x immunized against both infections.

Let $S_{00}^{post}(x)$ denote the probability that an individual of age x remains susceptible at time τ to both infections, $S_{01}^{post}(x)$ the probability of remaining susceptible to infection 1 but not infection 2, and so on. Then,

$$S_{00}^{post}(x) = \{1 - \pi_{12}(x)\}\left\{1 + \frac{\Lambda_1^{post}(x) + \Lambda_2^{post}(x)}{\theta}\right\}^{-\theta}$$

$$S_{01}^{post}(x) = \{1 - \pi_1(x)\}\left\{1 + \frac{\Lambda_1^{post}(x)}{\theta}\right\}^{-\theta} - S_{00}^{post}(x)$$

$$S_{10}^{post}(x) = \{1 - \pi_2(x)\}\left\{1 + \frac{\Lambda_2^{post}(x)}{\theta}\right\}^{-\theta} - S_{00}^{post}(x)$$

$$S_{11}^{post}(x) = 1 - S_{00}^{post}(x) - S_{01}^{post}(x) - S_{10}^{post}(x).$$

Note that, if the post-vaccination survey is conducted at time τ after the introduction of mass vaccination, then

$$\Lambda_i^{post}(x) = \begin{cases} \Lambda_i^{pre}(x - \tau) + \alpha_i(x) & x \geqslant \tau \\ \alpha_i(x) & x < \tau \end{cases}$$

where the $\alpha_i(x) \geqslant 0$. Suppose now that the post-vaccination survey yields n_{00x} individuals of age x susceptible to both infections, n_{01x} to infection 1 but not infection 2, etc. The kernel log-likelihood is

$$l^{post} = \sum_x \sum_{i,j} n_{ijx} \log\{S_{ij}^{post}(x)\}.$$

Given a suitable parametrization of the $\alpha_i(x)$, maximizing the joint log likelihood $l = l^{pre} + l^{post}$ yields estimates of $M(x)\beta_i^+(x, y)$, θ, and the $\alpha_i(x)$. The estimated value of $\overline{S}_i^{post}(x)$ is

$$\left\{1 + \frac{\widehat{\Lambda}_i^{post}(x)}{\widehat{\theta}}\right\}^{-(\widehat{\theta}+2)}$$

and the estimate of R_{ei} is

$$\widehat{R}_{ei} = (1 + \widehat{\theta}^{-1})\rho \left\{ M(x)\widehat{\beta}_i^+(x, y)\{1 - \pi_i(x)\}\left\{1 + \frac{\widehat{\Lambda}_i^{post}(x)}{\widehat{\theta}}\right\}^{-(\widehat{\theta}+2)}\right\}.$$

So far we have assumed that the quantities $\pi_1(x)$, $\pi_2(x)$ and $\pi_{12}(x)$ are known. In general, such information is required because there is no way of distinguishing between vaccine-derived and naturally induced immunity within each age group, and hence the $\pi_i(x)$ and the $\alpha_i(x)$ are not both identifiable.

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

However, in special cases it may be possible to estimate some of the $\pi_i(x)$. Such estimation is possible, for example, in the case of several infections with a common multivalent vaccine, since then the vaccine coverage is the same for all infections. Below we give a further example, where estimation of vaccine efficacy and catch-up MMR vaccine coverage is possible thanks to a parsimonious parameterisation of the $\alpha_i(x)$.

4.3 *A semi-parametric upper bound on R_e*

Both methods so far described require that the $\pi_i(x)$ be known. This requires accurate data on vaccine coverage and vaccine efficacy. If no such data are available, then the $\pi_i(x)$ are not known and hence the R_{ei} cannot be estimated. However, we can obtain useful upper bounds on semi-parametric estimates of the R_{ei}. We have

$$\begin{aligned}\{1-\pi_i(x)\}\overline{S}_i^{post}(x) &= P_i(x)\left\{\frac{P_i(x)}{1-\pi_i(x)}\right\}^{\frac{2}{\theta}}\\ &\leqslant P_i(x).\end{aligned}$$

This holds because $P_i(x) \leqslant 1-\pi_i(x)$ and $\theta > 0$. Thus

$$\begin{aligned}R_{ei} &= (1+\theta^{-1})\rho\left\{M(x)\{1-\pi_i(x)\}\overline{S}_i^{post}(x)\beta_i^+(x,y)\right\}\\ &\leqslant (1+\theta^{-1})\rho\left\{M(x)P_i(x)\beta_i^+(x,y)\right\}.\end{aligned}$$

This inequality follows from the following fact: if $A(x,y)$ and $B(x,y)$ are non-negative with non-negative leading eigenfunctions, and $B(x,y) = c(x)A(x,y)$ with $c(x) \geqslant 1$, then $\rho\{B(x,y)\} \geqslant \rho\{A(x,y)\}$.

The function $M(x)\beta_i^+(x,y)$ and the parameter θ are estimated from the first serological survey. The second survey provides non-parametric estimates of $P_i(x)$. Note that this procedure produces an upper bound for semi-parametric estimates only: parametric estimates might well exceed this upper bound.

5. APPLICATION TO MUMPS AND RUBELLA DATA

In this application we estimate R_e for mumps and rubella in 1996, $\tau = 8$ years after the introduction of MMR vaccine. We model $\beta_1^+(x,y)$ and $\beta_2^+(x,y)$ using 5×5 contact matrices with the same structure, and the standard age groups 0–5 years, 5–10 years, 10–15 years, 15–25 years, 25 + years. To illustrate the impact of model choice we use two matrix structures, which we label A and B. The two stuctures differ primarily in that A has a single isolated parameter β_2 describing contacts within the 5–10 year age group, whereas matrix B has two such parameters, β_2 and β_3, representing mixing within the 5–10 and the 10–15 year age groups (the values of the β_j will differ for mumps and rubella). The two structures are as follows:

$$A = \begin{pmatrix}\beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5\\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5\\ \beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5\\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5\\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5\end{pmatrix} \quad B = \begin{pmatrix}\beta_1 & \beta_1 & \beta_4 & \beta_4 & \beta_5\\ \beta_1 & \beta_2 & \beta_4 & \beta_4 & \beta_5\\ \beta_4 & \beta_4 & \beta_3 & \beta_4 & \beta_5\\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5\\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5\end{pmatrix}.$$

We shall assume a uniform age distribution on [0, 75], so that $m(x) = 75^{-1}$; results are insensitive to reasonable assumptions about age structure. Fitting these matrices to the 1987 paired serological survey data using the methods described by Farrington *et al.* (2001), assuming (a) no individual heterogeneity

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

Table 2. *R_0 for rubella and mumps from pre-vaccination survey, with bootstrap 95% confidence intervals*

	No individual heterogeneity		With individual heterogeneity	
	Rubella	Mumps	Rubella	Mumps
Matrix A	3.26	3.77	4.95	5.65
	(2.63, 5.66)	(3.57, 5.20)	(3.40, 9.59)	(4.57, 9.43)
Matrix B	3.36	10.61	5.06	19.70
	(2.67, 5.78)	(3.76, 18.16)	(3.55, 9.99)	(5.57, 50.21)

Table 3. *R_e for rubella and mumps: semi-parametric upper bounds and central estimates, with bootstrap 95% confidence intervals*

	No individual heterogeneity		With individual heterogeneity			
			Upper bounds		Central estimates	
	Rubella	Mumps	Rubella	Mumps	Rubella	Mumps
Matrix A	0.27	0.67	0.40	0.99	0.27	0.87
	(0.20, 0.44)	(0.52, 0.86)	(0.26, 0.78)	(0.70, 1.40)	(0.19, 0.49)	(0.60, 1.18)
Matrix B	0.27	1.49	0.41	2.76	0.28	1.93
	(0.21, 0.47)	(0.58, 2.88)	(0.28, 0.81)	(0.87, 7.73)	(0.20, 0.52)	(0.73, 5.17)

and (b) individual heterogeneity with a gamma distribution, gives the estimates of R_0 shown in Table 2. The log-likelihood ratio statistic for individual heterogeneity, common to both matrices, is 15.42 on one degree of freedom, $p < 0.001$. The estimated frailty variance is $\widehat{\theta}^{-1} = 0.1336$. Note that the values of R_0 with individual heterogeneity are greater than $1 + \widehat{\theta}^{-1}$ times those without heterogenity. This is because the estimates of $M(x)\beta^{+}(x, y)$ change: if there is heterogeneity but the model is fitted assuming none, the estimated matrix is biased. The age-specific proportions immunised in 1996 were calculated as follows. Let $i = 1$ denote mumps, and $i = 2$ denote rubella. Thus, for mumps, $\pi_1(x) = C_{MMR}(1996 - x) \times 0.85$, where $C_{MMR}(t)$ is the MMR vaccine coverage achieved for children born in year t and 0.85 is the mumps vaccine efficacy. For rubella, $\pi_2(x) = 1 - (1 - C_{MMR}(1996 - x) \times 0.95) \times (1 - C_{MR}(1996 - x) \times 0.95)$, where $C_{MR}(t)$ is the MR vaccine coverage and 0.95 is the rubella vaccine efficacy. The assumptions about vaccine coverage and efficacy were described in Section 2. For the parametric model, we also need $\pi_{12}(x)$. Assuming that both the uptake of MMR and MR vaccines, and the responses to the mumps and rubella components of MMR are independent within individuals, this is $1 - (1 - C_{MMR}(1996 - x) \times 0.95 \times 0.85) \times (1 - C_{MR}(1996 - x) \times 0.95)$.

Table 3 shows the estimates of R_e in 1996, both with and without individual heterogeneity. For the estimates with individual heterogeneity, Table 3 gives the semi-parametric upper bounds and central estimates, with confidence intervals. The parametric estimates are shown in Table 4. These were obtained by assuming that, after the introduction of MMR in 1988, the force of infection was reduced to some low constant value. Thus

$$\alpha_i(x) = \begin{cases} \alpha_i x & x < \tau \\ \alpha_i \tau & x \geqslant \tau. \end{cases}$$

The first two columns of Table 4 shows the estimates obtained using an assumed 'catch-up' MMR vaccine coverage of 60% in the 1984–86 birth cohorts and assumed vaccine efficacies of 85% for mumps and 95% for rubella. Using this model, $\widehat{\theta}^{-1} = 0.1669$, $\widehat{\alpha}_1 = 0.0469$ (mumps) and $\widehat{\alpha}_2 = 0.0198$ (rubella). We also fitted a model using the published MMR vaccine coverage from 1987, but with free parameters for

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

Table 4. *R_e for rubella and mumps: parametric estimates, with profile likelihood 95% confidence intervals*

	With assumed efficacies and catch-up coverage		With estimated efficacies and catch-up coverage	
	Rubella	Mumps	Rubella	Mumps
Matrix A	0.28	1.01	0.29	1.04
	(0.19, 0.46)	(0.76, 1.38)	(0.19, 0.47)	(0.78, 1.43)
Matrix B	0.29	3.32	0.29	3.16
	(0.20, 0.47)	(0.80, 7.19)	(0.19, 0.47)	(0.81, 7.08)

the mumps and rubella vaccine efficacies and for the average MMR vaccine catch-up campaign coverage in the 1984–86 birth cohorts. This produced a slight reduction in the deviance of 4.363. The values of R_e obtained with this model are shown in columns 3 and 4 of Table 4. The estimated vaccine efficacies were 0.851 for mumps and 0.977 for rubella. The catch-up coverage was estimated to be 73.9%, rather higher than the 60% figure we had assumed. We also obtained $\widehat{\theta}^{-1} = 0.1642$, $\widehat{\alpha}_1 = 0.0379$ (mumps) and $\widehat{\alpha}_2 = 0.0000$ (rubella).

Figure 1 shows the observed and fitted serological profiles in 1987 and 1996 using this parametric model. The irregularities in the fitted seroprevalence in the 1996 survey reflect the changes in vaccine coverage. The published vaccination coverage data may be inaccurate, and further exploration of coverage levels using parametric modelling might be possible. Our assumption that post-MMR infection rates are constant is probably simplistic, though we doubt that relaxing it would make much difference since most older people are immune. Overall, we recommend using either the semi-parametric upper bounds on R_e, or the parametric estimates incorporating estimation of any unknown vaccine coverage or vaccine efficacy. For the present data, these methods produce broadly similar results.

The confidence intervals for R_0 and the semi-parametric upper bounds and central estimates of R_e are bootstrap percentile intervals based on 599 samples. Bootstrapping proved too time-consuming for the parametric models for R_e. For these we used profile likelihood 95% confidence intervals, which we calculated using the method of Critchley *et al.* (1988).

Individual heterogeneity, as expected, has a big impact on R_e, particularly for mumps, though less than its impact on R_0. Ignoring individual heterogeneity produces estimates that are too low. As with R_0, estimates of R_e are very sensitive to the choice of model matrix; it was suggested in Farrington *et al.* (2001) that the evidence favoured matrix B. The conclusion from this analysis is that rubella was well-controlled in 1996 in the UK, but that mumps may not have been.

This interpretation is supported by surveillance data, obtained from the PHLS website. In 1996 there were 94 laboratory confirmed cases of mumps in England and Wales, rising steadily to 759 in 2001, in spite of the introduction of a second dose of MMR vaccine. Notifications, both confirmed and unconfirmed, show a similar trend. In contrast, rubella outbreaks in universities in 1996 did not lead to substantial spread in the wider community (Miller *et al.*, 1997), and the number of cases has declined steadily since 1996. These observations are consistent with $R_e < 1$ for rubella and $R_e > 1$ for mumps in the period after 1996.

6. DISCUSSION

In this paper we have extended methods for estimating the effective reproduction number R_e from age-stratified serological survey data to include the effect of individual heterogeneity. As illustrated by the mumps and rubella example, individual heterogeneity can have a substantial effect on the estimates of

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

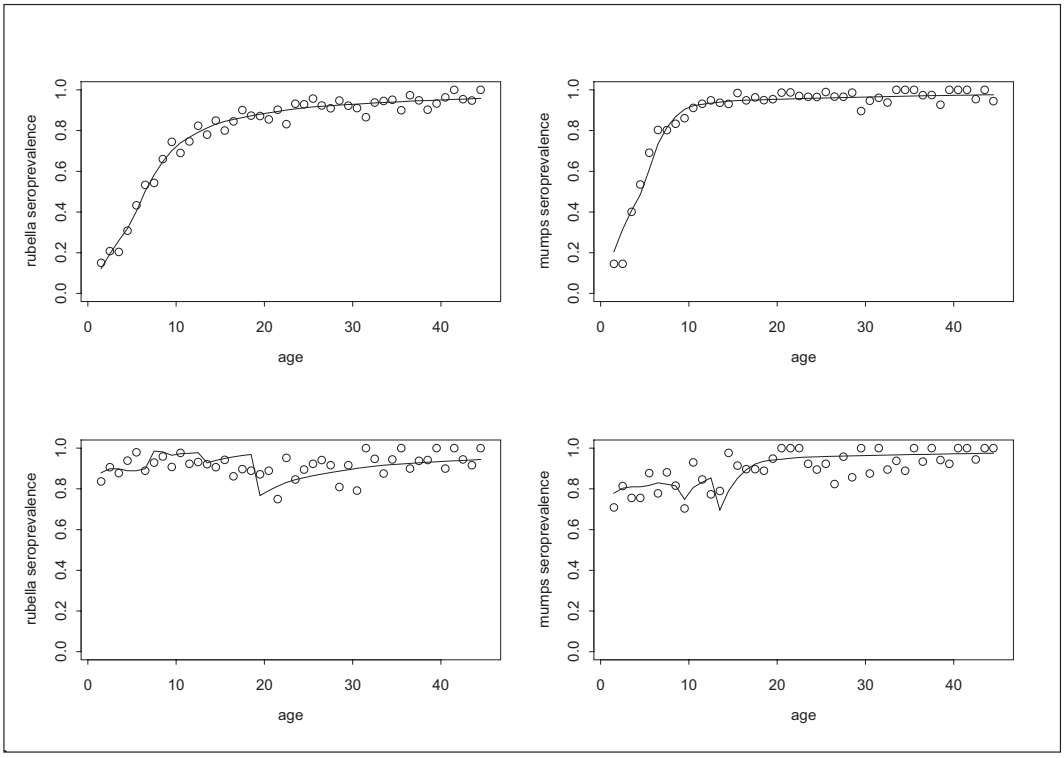

Fig. 1. Observed and fitted serological profiles. Top: 1987 survey, bottom: 1996 survey. Left: rubella, right: mumps.

both R_0 and R_e. Allowing for individual heterogeneity, however, calls for greater modelling complexity.

The parametric models we described require paired data from two surveys, conducted before and after the introduction of mass vaccination. In view of the non-identifiability of vaccine-induced and naturally acquired immunity, accurate data on vaccine coverage and vaccine efficacy are also generally required. Ideally, serological surveys should be supplemented by vaccine coverage surveys, which could be carried out relatively easily using vaccination databases. In some circumstances, however, some of the parameters describing vaccine-induced immunity levels can be estimated, including vaccine efficacy and coverage, as illustrated in our example. Thus, the parametric model makes optimal use of all the information available.

If paired data are not available in the post-vaccination survey, then a semi-parametric estimate of R_e may be obtained. We used the raw observed proportions susceptible $P(x)$, adjusted to avoid conflict with the immunization data; it might be better to smooth the data a little. The semi-parametric method, however, still requires information on age-specific immunisation rates.

If such data are unavailable, or of dubious quality, then we propose an upper bound on semi-parametric estimates of the effective reproduction number, which can be calculated without any assumptions about vaccination coverage or efficacy. In the absence of individual heterogeneity, this upper bound and the semi-parametric estimate are identical.

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022

ACKNOWLEDGEMENTS

This work was supported by Wellcome Trust project grant 061830. We thank Elizabeth Miller and Nigel Gay of the Communicable Disease Surveillance Centre, London, for making the serological survey data available to us, and an anonymous referee for helpful comments.

REFERENCES

CRITCHLEY, F., FORD, I. AND RIJAL, O. (1988). Interval estimation based on the profile likelihood: Strong Lagrangian theory with applications to discrimination. *Biometrika* **75**, 21–28.

DIEKMANN, O., HEESTERBEEK, J. AND METZ, J. (1990). 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* **28**, 365–382.

FARRINGTON, C. P. (2003). On vaccine efficacy and reproduction numbers.Accepted for publication in *Mathematical Biosciences*.

FARRINGTON, C. P., KANAAN, M. N. AND GAY, N. J. (2001). Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics* **50**, 1–33.

FARRINGTON, C. P., KANAAN, M. N. AND GAY, N. J. (2003). Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* **4**, 279–295.

GAY, N. J., HESKETH, L. M., MORGAN-CAPNER, P. AND MILLER, E. (1995). Interpretation of serological surveillance data for measles using mathematical models: Implications for vaccine strategy. *Epidemiology and Infection* **115**, 139–156.

MILLER, E., WAIGHT, P., GAY, N., RAMSAY, M., VURDIEN, J., MORGAN-CAPNER, P., HESKETH, L., BROWN, D., TOOKEY, P. AND PECKHAM, C. (1997). The epidemiology of rubella in England and Wales before and after the 1994 measles and rubella vaccination campaign: Fourth joint report from the PHLS and the national congenital rubella surveillance programme. *Communicable Disease Report Review* **7**, R26–R32.

WALLINGA, J., LEVY-BRUHL, D., GAY, N. J. AND WACHMANN, C. H. (2001). Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiology and Infection* **127**, 281–295.

WHITAKER, H. J. AND FARRINGTON, C. P. *Estimation of infectious disease parameters from serological survey data: the impact of regular epidemics*. submitted.

WIRATUNGA, E. B. P. AND O'BRIEN, J. (1995). National measles and rubella vaccination campaign. *British Medical Journal* **310**, 1532.

[*Received November 11, 2002; revised February 4, 2003; accepted for publication February 10, 2003*]

Downloaded from https://academic.oup.com/biostatistics/article/4/4/621/246942 by guest on 20 August 2022