Epidemiol. Infect. (2009), **137**, 48–57. © 2008 Cambridge University Press
doi:10.1017/S0950268808000563 Printed in the United Kingdom

Contact patterns and their implied basic reproductive numbers: an illustration for varicella-zoster virus

T. VAN EFFELTERRE[1]*, Z. SHKEDY[1], M. AERTS[1], G. MOLENBERGHS[1], P. VAN DAMME[2] AND P. BEUTELS[2,3]

[1] *Hasselt University, Center for Statistics, Biostatistics, Diepenbeek, Belgium*
[2] *University of Antwerp, Epidemiology and Social Medicine, Center for Evaluation of Vaccination, Antwerp, Belgium*
[3] *National Centre for Immunization Research and Surveillance, University of Sydney, Australia*

(*Accepted 25 February 2008; first published online 9 May 2008*)

SUMMARY

The WAIFW matrix (Who Acquires Infection From Whom) is a central parameter in modelling the spread of infectious diseases. The calculation of the basic reproductive number (R_0) depends on the assumptions made about the transmission within and between age groups through the structure of the WAIFW matrix and different structures might lead to different estimates for R_0 and hence different estimates for the minimal immunization coverage needed for the elimination of the infection in the population. In this paper, we estimate R_0 for varicella in Belgium. The force of infection is estimated from seroprevalence data using fractional polynomials and we show how the estimate of R_0 is heavily influenced by the structure of the WAIFW matrix.

INTRODUCTION

An essential assumption in modelling the spread of infectious diseases is that the force of infection, which is the probability for a susceptible to acquire the infection, varies over time as a function of the level of infectivity in the population [1]. For many infectious diseases, the force of infection is also known to depend on age. The equation describing the dependence of the force of infection on age and time is given by

$$\lambda(a,t)=\int_0^L \beta(a,a')I(a',t)\mathrm{d}a'. \quad (1)$$

The coefficients $\beta(a, a')$ are called the transmission coefficients and $I(a', t)$ is the number of infectious individuals at age a' and time t. These transmission coefficients combine epidemiological, environmental and social factors affecting the transmission rate between an infective of age a' and a susceptible of age a [1, 2]. For the discrete case with a population divided into a finite number, say n, of age groups, Anderson & May [3] introduced the WAIFW (Who Acquires Infection From Whom) matrix in which the ij th entry of the matrix, β_{ij}, is the transmission coefficient from an infective in age group j to a susceptible in age group i. Let $\bar{I}_i$ be the total number of infectious individuals in the ith age group at time t, $i=1, \ldots, n$, then the age- and time-dependent force of infection can be approximated by the matrix product

$$\lambda = W\bar{I}. \quad (2)$$

Here, $\bar{I}=(\bar{I}_1, \ldots, \bar{I}_n)$ is the vector in which the ith element is the number of infectious individuals (prevalence of infectivity) in age group i, $\lambda=(\lambda_1, \ldots, \lambda_n)$ is the vector in which the ith element is the force of infection specific to age group i and W is a known WAIFW matrix. The configuration of the WAIFW matrix represents *a priori* knowledge (or assumptions)

* Author for correspondence: Dr T. Van Effelterre, Hasselt University, Center for Statistics, Biostatistics, Agoralaan 1, B3590 Diepenbeek, Belgium.
(Email: tvaneff@yahoo.com)

about the mixing patterns in the population. Several configurations are discussed in the literature (see e.g. [1, 2, 4–6]). For example, for a model with five age groups the WAIFW matrix W_1 in equation (3) represents a mixing pattern for which individuals are mixing only with individuals from their own age group (assortative mixing [5]) with a specific age-dependent transmission coefficient while W_2 represents a mixing pattern similar to W_1, also accounting for an additional mixing of individuals with individuals of other age groups with a 'background' transmission coefficient:

$$W_1 = \begin{pmatrix} \beta_1 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 \\ 0 & 0 & 0 & 0 & \beta_5 \end{pmatrix}, \quad W_2 = \begin{pmatrix} \beta_1 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_2 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_3 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}. \tag{3}$$

Note that both matrices have five unknown parameters and both are symmetric. For each of these contact structures, the number of parameters is equal to the number of age groups. This is a condition to have a solution [3]. Suppose that the population is divided into n age groups and let $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$ be the estimated vector of age-specific force of infection in each age group. If the structure of the WAIFW matrix is known and consists of n unknown parameters, the WAIFW matrix can be estimated using the equality

$$\begin{pmatrix} \hat{\lambda}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\lambda}_n \end{pmatrix} = \frac{ND}{L} W \begin{pmatrix} \Psi_1 \\ \cdot \\ \cdot \\ \cdot \\ \Psi_n \end{pmatrix}, \tag{4}$$

with N the total population size, D the mean duration of infectiousness, L the life-expectancy at birth, and

$$\Psi_j = e^{-\varphi_{j-1}} - e^{-\varphi_j} \quad \text{and} \quad \varphi_j = \sum_{i=1}^{j} \hat{\lambda}_i (a_i - a_{i-1}). \tag{5}$$

Here, $a_i - a_{i-1}$ is the width of the ith age group. Hence, as long as the WAIFW matrix has a known configuration with n unknown parameters, the parameter vector $\beta = (\beta_1, \ldots, \beta_n)$ is identifiable. Note that we expect that $\beta_i \geqslant 0$, $i = 1, 2, \ldots, n$.

The basic reproductive number R_0 can be computed as the dominant eigenvalue of a matrix for which the ijth entry is the basic reproductive number $R_{0_{ij}}$, specific to the transmission from an infective in age group j to a susceptible in age group i. More precisely, $R_{0_{ij}} = \beta_{i,j} D N_i$ where D is the duration of infectiousness assumed independent of age and N_i is the size of the population in age group i. Therefore, the estimator for R_0 depends on the configuration of the WAIFW matrix. Farrington *et al.* [4] showed that different configurations of the WAIFW matrix can lead to quite different estimates for R_0. For example, Farrington *et al.* [4] estimated R_0 for mumps to be equal to 25·5, 8·0 and 3·3 for the configuration of W_2, W_3 and W_4, respectively:

$$W_3 = \begin{pmatrix} \beta_1 & \beta_1 & \beta_4 & \beta_4 & \beta_5 \\ \beta_1 & \beta_2 & \beta_4 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}, \quad W_4 = \begin{pmatrix} \beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ \beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}. \tag{6}$$

Hence, the uncertainty related to the WAIFW matrix is coming from two different sources: (1) the uncertainty about the unknown transmission coefficients β_i and (2) the uncertainty about the configuration of the WAIFW matrix. Furthermore, Wallinga *et al.* [6] showed that the basic reproductive number for measles ranges between 770·38 when assortative mixing pattern is assumed and 1·43 when infant mixing is assumed, i.e. infants are assumed to be the source of all infection [6].

In this paper, we present an investigation of the estimation of the basic reproductive number, R_0, and p_c, the minimal proportion of the population that needs to be vaccinated to eliminate the infection, for varicella in Belgium for which, currently, there is no vaccination programme. Following the approach of Greenhalgh & Dietz [5] we show that, depending on our assumption about the contact patterns, R_0 ranges between 3·12 and 68·57, and p_c ranges between 67·9% and 98·5%.

This paper is organized as follow. In the next section, we present six possible configurations for the WAIFW

matrix for varicella and discuss the estimation of the age-dependent force of infection from serological data using fractional polynomials. In the 'Estimation of the transmission coefficients' section, the WAIFW matrices are estimated using the integrated force of infection in the relevant age groups. Parametric and non-parametric bootstrap is used to calculate confidence intervals for the transmission coefficients. The estimation of R_0 and p_c is discussed in the 'Estimation of R_0 from the WAIFW matrices' section.

Estimation of R_0 and the WAIFW matrix for varicella based on serological data

Age-dependent transmission coefficients

Several authors [4, 5, 7] illustrated how the estimation of R_0 is influenced by the configuration of the WAIFW matrix. A specific configuration of the WAIFW matrix represents specific assumptions about age-dependent transmission coefficients in the population which in turn represents prior assumptions about the mixing patterns in the population.

We illustrate these concepts for varicella in Belgium. The population was divided into the following six age groups, taking into account the schooling system in Belgium, 6 months–1 year, 2–5 years, 6–11 years, 12–18 years, 19–30 years and 31–44 years. Several configurations were discussed by Anderson & May [1] and Greenhalgh & Dietz [5]. Assortative mixing assumes that all contacts occurs within the age groups [5]. The matrix W_{V1} has specific transmission coefficients within each age group, i.e. for the transmission among hosts belonging to the same age group on the diagonal, and of a common 'background' transmission coefficient between any two different age groups. This 'background' transmission coefficient is assumed equal to the transmission coefficients in the oldest age group (β_6). Note that the transmission coefficient in the oldest age group and the 'background' transmission coefficient between different age groups are both expected to be smaller than the transmission coefficients in younger age groups. The second configuration, W_{V2}, assumes that the main route of transmission for a directly transmitted viral infection like varicella is in kindergarten children or in the classroom. This is expressed by a unique coefficient β_2 for the (presumed high) transmission between infectious and susceptible hosts in the range 2–5 years and two other specific coefficients (β_1 and β_3) for transmission amongst other hosts aged <12 years. The third configuration of the WAIFW matrix, W_{V3} is a minor variation of the structure W_{V2} with a common transmission coefficient β_1 between hosts in the range 6 months–1 year and hosts aged <12 years. For the fourth configuration, W_{V4}, the transmission coefficient depends only on the age group of the susceptible host. The susceptible hosts of a given age group are assumed to be as likely to acquire infection from infectious hosts of any age. Note that this structure is not symmetric. The fifth matrix structure, W_{V5}, is a minor variation of the assortative structure W_{V1} in which there is a common transmission coefficient in the two older age groups and a distinct 'background' coefficient for the transmission between any other two different age groups. Finally, the sixth configuration, W_{V6}, is the extreme case of assortative mixing for which there is only transmission assumed between hosts belonging to the same age group. Of these proposed configurations, the latter structure is most obviously unrealistic, it is nevertheless useful as it provides an upper bound to the basic reproductive number R_0 [5].

$$W_{V1}=\begin{pmatrix}\beta_1&\beta_6&\beta_6&\beta_6&\beta_6&\beta_6\\ \beta_6&\beta_2&\beta_6&\beta_6&\beta_6&\beta_6\\ \beta_6&\beta_6&\beta_3&\beta_6&\beta_6&\beta_6\\ \beta_6&\beta_6&\beta_6&\beta_4&\beta_6&\beta_6\\ \beta_6&\beta_6&\beta_6&\beta_6&\beta_5&\beta_6\\ \beta_6&\beta_6&\beta_6&\beta_6&\beta_6&\beta_6\end{pmatrix},\quad W_{V2}=\begin{pmatrix}\beta_1&\beta_1&\beta_3&\beta_4&\beta_5&\beta_6\\ \beta_1&\beta_2&\beta_3&\beta_4&\beta_5&\beta_6\\ \beta_3&\beta_3&\beta_3&\beta_4&\beta_5&\beta_6\\ \beta_4&\beta_4&\beta_4&\beta_4&\beta_5&\beta_6\\ \beta_5&\beta_5&\beta_5&\beta_5&\beta_5&\beta_6\\ \beta_6&\beta_6&\beta_6&\beta_6&\beta_6&\beta_6\end{pmatrix},$$

$$W_{V3}=\begin{pmatrix}\beta_1&\beta_1&\beta_1&\beta_4&\beta_5&\beta_6\\ \beta_1&\beta_2&\beta_3&\beta_4&\beta_5&\beta_6\\ \beta_1&\beta_3&\beta_3&\beta_4&\beta_5&\beta_6\\ \beta_4&\beta_4&\beta_4&\beta_4&\beta_5&\beta_6\\ \beta_5&\beta_5&\beta_5&\beta_5&\beta_5&\beta_6\\ \beta_6&\beta_6&\beta_6&\beta_6&\beta_6&\beta_6\end{pmatrix},\quad W_{V4}=\begin{pmatrix}\beta_1&\beta_1&\beta_1&\beta_1&\beta_1&\beta_1\\ \beta_2&\beta_2&\beta_2&\beta_2&\beta_2&\beta_2\\ \beta_3&\beta_3&\beta_3&\beta_3&\beta_3&\beta_3\\ \beta_4&\beta_4&\beta_4&\beta_4&\beta_4&\beta_4\\ \beta_5&\beta_5&\beta_5&\beta_5&\beta_5&\beta_5\\ \beta_6&\beta_6&\beta_6&\beta_6&\beta_6&\beta_6\end{pmatrix},$$

$$W_{V5}=\begin{pmatrix}\beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6\\ \beta_6 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6\\ \beta_6 & \beta_6 & \beta_3 & \beta_6 & \beta_6 & \beta_6\\ \beta_6 & \beta_6 & \beta_6 & \beta_4 & \beta_6 & \beta_6\\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 & \beta_6\\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5\end{pmatrix}, \quad W_{V6}=\begin{pmatrix}\beta_1 & 0 & 0 & 0 & 0 & 0\\ 0 & \beta_2 & 0 & 0 & 0 & 0\\ 0 & 0 & \beta_3 & 0 & 0 & 0\\ 0 & 0 & 0 & \beta_4 & 0 & 0\\ 0 & 0 & 0 & 0 & \beta_5 & 0\\ 0 & 0 & 0 & 0 & 0 & \beta_6\end{pmatrix}.$$

Estimation of the age-dependent force of infection for varicella

The estimation of the WAIFW matrix requires the estimation of the force of infection from pre-vaccination data. Following the methodology proposed by Anderson & May [1] (see also [2, 7]) we assume that an age-specific serological profile can be estimated from pre-vaccination data. This can be done by modelling pre-vaccination seroprevalence data. The age-dependent force of infection can be derived from the estimated model for the prevalence of seropositive hosts by using non-parametric methods (discussed in [8, 9]), or parametric methods (discussed in [4, 10–13]).

For varicella, we estimated the force of infection by using a seroprevalence dataset consisting of 1673 individuals aged between 1 and 44 years that was sampled in Antwerp (Belgium) between October 1999 and April 2000 and reported by Thiry *et al.* [14]. The sera were residual specimens submitted to medical laboratories for diagnostic purposes. Sera for the 1–11 years age group were collected from outpatients hospitals in Antwerp, sera for the 12–18 years age group were collected from volunteers in vaccine trials and sera for age groups >16 years were provided by a medical laboratory in Antwerp. The population was stratified by age in order to sample about 100 observations per age group. The force of infection can be estimated from this serological sample under the assumption that the disease is in a steady state.

For the analysis presented in this paper, fractional polynomial [15] models were used to describe the dependency of the force of infection on age, as discussed in Shkedy *et al.* [13]. Briefly, a generalized linear model for the binary data with logit link was used to estimate the force of infection. The linear predictor for that model is given by

$$\eta_m(a,\beta,p_1,p_2,\ldots,p_m)=\sum_{i=0}^{m}\beta_i H_i(a), \tag{7}$$

where m is an integer, $p_1<p_2<\ldots<p_m$ is a sequence of powers and $H_i(a)$ is a transformation function given by

$$H_i(a)=\begin{cases} a^{p_i} & \text{if } p_i\neq p_{i-1},\\ H_{i-1}(a)\times\log(a) & \text{if } p_i=p_{i-1},\end{cases} \tag{8}$$

with $p_0=0$ and $H_0=1$. As shown in Shkedy *et al.* [13] the force of infection in this model can be expressed as

$$\lambda(a)=\eta_m(a,\beta,p_1,p_2,\ldots,p_m)'\frac{e^{\eta_m(a,\beta,p_1,p_2,\ldots,p_m)}}{1+e^{\eta_m(a,\beta,p_1,p_2,\ldots,p_m)}}, \tag{9}$$

where $\eta_m(a,\beta,p_1,p_2,\ldots,p_m)'$ denotes the partial derivative of $\eta_m(a,\beta,p_1,p_2,\ldots,p_m)$ with respect to age a. Figure 1 shows, for the varicella dataset, the estimated model for the prevalence of seropositive hosts (Fig. 1*a*) and the force of infection (Fig. 1*b*). Constrained fractional polynomials were fitted to ensure that the estimated force of infection will be non-negative. The model fit was based on the value of the Akaike Information Criterion (AIC) and the selected model has exponents $p_1=-0{\cdot}4$ and $p_2=-0{\cdot}3$ with an AIC equal to 125·769.

The estimated force of infection is given by equation (9), with $m=2$ and $\eta_2(a)=-40{\cdot}231\,a^{-0{\cdot}3}+28{\cdot}153\,a^{-0{\cdot}4}+11{\cdot}303$. According to this model, the force of infection for varicella in Belgium peaks at 2 years of age with a value of $\ell(2)=0{\cdot}3111$ and drops monotonically for older susceptibles. At 44 years of age the force of infection is estimated to be 0·0315. The mean force of infection for each of the six age groups can be estimated by integrating from the parametric model of the force of infection in equation (9). Hence, we first use a flexible parametric model to estimate the force of infection and integrate over the age groups thereafter. The advantage of using fractional polynomials is that this integration can be performed analytically and the force of infection in age group i ($i=1,\ldots,6$) is given by

$$\lambda_i=\frac{\log\left(\frac{1+e^{\eta_2(a_i)}}{1+e^{\eta_2(a_{i-1})}}\right)}{a_i-a_{i-1}}, \tag{10}$$

where a_{i-1} and a_i are the lower and upper bounds of age group i, respectively. The estimates for the six age groups were $\hat{\lambda}_1=0{\cdot}254$, $\hat{\lambda}_2=0{\cdot}267$, $\hat{\lambda}_3=0{\cdot}160$,

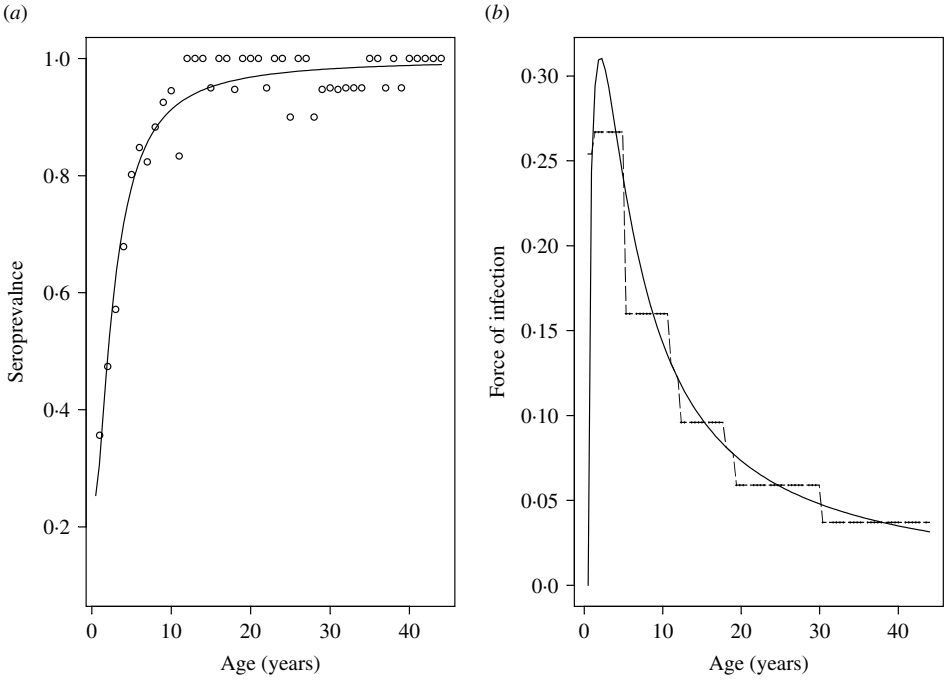

Fig. 1. Estimated prevalence (*a*) and force of infection (*b*) for varicella in Belgium. – – –, Integrated force of infection.

$\hat{\lambda}_4 = 0{\cdot}096$, $\hat{\lambda}_5 = 0{\cdot}059$ and $\hat{\lambda}_6 = 0{\cdot}037$ for age groups 6 months–1 year, 2–5 years, 6–11 years, 12–18 years, 19–30 years and 31–44 years, respectively. Figure 1*b* shows the force of infection estimated by the fractional polynomial (solid line) and the integrated force of infection (dashed line) for the six age groups.

Estimation of the transmission coefficients

Once the estimates of the force of infection are obtained, the elements of the WAIFW matrix can be computed by

$$\frac{L}{D \times N}\hat{\lambda}_i = \sum_{j=1}^{n} \beta_{ij}\Psi_j,$$

where L is the life-expectancy at birth, D is the mean duration of infectiousness, N is the total population size and Ψ_j is given by equation (5). Substituting the $\hat{\lambda}_i$s by their expression, we have

$$e^{-\varphi_0} = 1, \tag{11}$$

$$e^{-\varphi_1} = \frac{1 + e^{\eta(a_0)}}{1 + e^{\eta(a_1)}}. \tag{12}$$

It is easy to show that for $i = 2, \ldots, 6$:

$$e^{-\varphi_i} = \Pi_{j=1}^{i} \frac{1 + e^{\eta(a_{j-1})}}{1 + e^{\eta(a_j)}}. \tag{13}$$

Hence, in our model for varicella using six age groups, the expressions for the Ψ_i s are

$$\Psi_i = e^{-\varphi_{i-1}} - e^{-\varphi_i} = \frac{(1 + e^{\eta(a_0)})(e^{\eta(a_i)} - e^{\eta(a_{i-1})})}{(1 + e^{\eta(a_{i-1})})(1 + e^{\eta(a_i)})}, \tag{14}$$

for $i = 1, \ldots, 6$, where $a_0 = 0{\cdot}5$, $a_1 = 2$, $a_2 = 6$, $a_3 = 12$, $a_4 = 19$, $a_5 = 31$, $a_6 = 45$, $N = 10237988$, $D = 7/365$ years and $L = 78$ years.

System (4) ($i = 1, \ldots, n$) is a linear system of n equations in n^2 unknowns: the elements of the WAIFW matrix β_{ij} ($i, j = 1, \ldots, n$). Since this system is underdetermined, we need to impose a structure upon the WAIFW matrix, limiting the number of unknowns to n ($n = 6$ for our model of varicella). We have estimated the elements of the WAIFW matrix for the six types of matrix structure described above. For example, the estimate of β_6 for the WAIFW matrix structure W_{V3} is given by

$$\beta_6 = \frac{\frac{L}{DN}\hat{\lambda}_6}{\Psi_1 + \Psi_2 + \Psi_3 + \Psi_4 + \Psi_5 + \Psi_6}. \tag{15}$$

The expressions for the other β_is are given in the Appendix for matrix W_{V3}. Similar expressions can be derived for each of the five other matrix structures.

The estimates of the β_is are given in Table 1 for the six WAIFW matrix structures described above (see also Fig. 2), together with two types of confidence intervals (CIs): the non-parametric bootstrap percentile 95%

Table 1. *Parameter estimates for the transmission coefficients. Decimal points are shown to illustrate differences in the confidence intervals*

Matrix	Parameter	Estimate ($\times 10^{-5}$)	95% confidence intervals ($\times 10^{-5}$)	
			Non-parametric	Parametric
W_{V1}	β_1	28·6682	27·7358–29·5323	27·5719–29·5306
	β_2	21·8524	20·3157–23·6582	20·4073–23·7245
	β_3	35·1168	29·0089–43·0572	29·8378–43·4135
	β_4	54·5450	40·7348–75·4701	42·3997–76·1505
	β_5	38·7205	26·5479–58·3102	28·0602–59·4965
	β_6	1·5094	1·3360–1·7058	1·3609–1·7053
W_{V2}	β_1	11·6105	10·4412–12·7368	10·4131–12·7302
	β_2	12·7365	10·8082–14·8602	10·8940–14·6658
	β_3	6·7183	5·9755–7·6053	6·0147–7·5431
	β_4	3·9368	3·4821–4·4661	3·5077–4·4259
	β_5	2·3970	2·1192–2·7169	2·1350–2·6917
	β_6	1·5094	1·3351–1·7081	1·3459–1·6923
W_{V3}	β_1	10·8318	9·9369–11·7734	9·9768–11·7272
	β_2	13·9985	12·5165–15·6485	12·5210–15·5900
	β_3	4·5187	2·9861–6·0658	2·9490–6·0417
	β_4	3·9368	3·4693–4·4508	3·4827–4·4379
	β_5	2·3970	2·1111–2·7083	2·1205–2·6978
	β_6	1·5094	1·3306–1·7022	1·3371–1·6951
W_{V4}	β_1	10·2380	9·3540–11·2279	9·3234–11·1187
	β_2	10·7494	9·6824–12·0692	9·6734–11·9518
	β_3	6·4435	5·6806–7·4510	5·6677–7·2881
	β_4	3·8779	3·4311–4·4542	3·4111–4·3827
	β_5	2·3887	2·1112–2·7300	2·1051–2·6929
	β_6	1·5094	1·3364–1·7188	1·3334–1·6960
W_{V5}	β_1	29·8724	28·9071–30·6015	28·9620–30·6812
	β_2	22·5377	21·0784–24·2330	21·0108–24·3728
	β_3	38·4309	32·1284–46·5073	32·2774–46·5480
	β_4	66·7451	50·7934–89·8739	51·2996–90·4674
	β_5	62·2835	44·8055–89·4235	45·6252–90·3289
	β_6	0·9391	0·7546–1·1525	0·7657–1·1486
W_{V6}	β_1	31·8553	31·3491–32·3493	31·3605–32·3941
	β_2	23·6662	22·130–25·2923	22·0467–25·6191
	β_3	43·8884	36·7338–53·0386	36·4274–54·2846
	β_4	86·8345	65·136–117·256	64·743–123·908
	β_5	101·083	69·503–150·779	69·123–160·334
	β_6	162·356	103·335–263·287	102·905–280·916

CIs and the parametric bootstrap percentile 95% CIs. We now provide details on the computations.

Bootstrap confidence intervals

In the seroprevalence sample, we have 44 samples by age of size N_i ($i=1, \ldots, 44$) and let p_i be the proportion of subjects of age i who are seropositive for varicella-zoster virus antibodies. The total number of subjects in the seroprevalence sample is $N=\sum N_i$. A non-parametric bootstrap sample is a sample of size N obtained by drawing a random sample (with replacement) for each age i. The sample for age i is a sample of size N_i drawn from a Bernoulli distribution with probability p_i. Since we are interested in the number of subjects seropositive for varicella-zoster virus in the sample, we can equivalently generate a random value from the binomial distribution (N_i, p_i) for each age i. We have used 1000 bootstrap samples in the computation since this is deemed necessary for the estimation of confidence intervals. For each bootstrap sample, we can then estimate the parameters of the model. For these estimates, we can estimate the mean force of infection $\lambda_1, \lambda_2, \ldots, \lambda_6$, in each of the six age groups. In this way we obtain 1000 bootstrap values

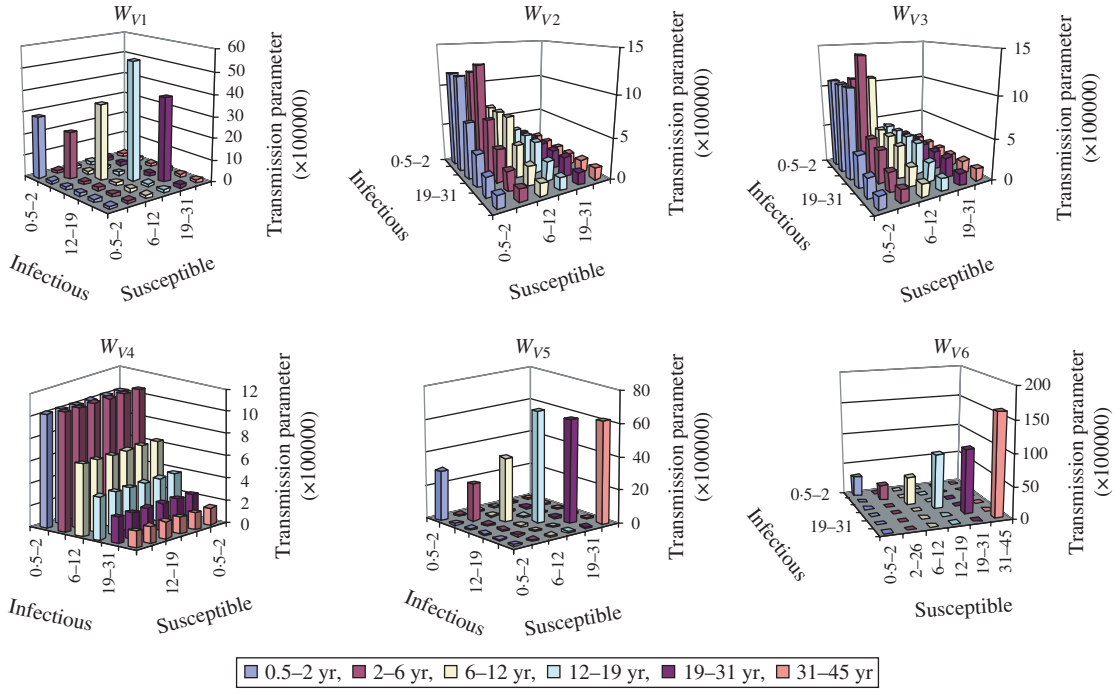

Fig. 2. WAIFW matrices for varicella in Belgium. Configurations 1–6. The upper limit is not included in the category age.

for each λ_i ($i=1, \ldots, 6$). The non-parametric percentile bootstrap 95% CI for a parameter (for example λ_1), is obtained from the bootstrap distribution of λ_1 by taking the 25th and the 975th value in the sequentially ordered set of 1000 bootstrap values of this parameter.

The parametric bootstrap and the computation of the parametric percentile bootstrap 95% CIs is performed in the same way as described above for the non-parametric bootstrap except that instead of drawing for each age i a random sample with replacement from the Bernoulli distribution with probability p_i, i.e. from the data, we can use instead the parametric model for the cumulative distribution of the age at infection $\pi(a)$ and draw the random sample from the Bernoulli distribution with probability f_i, where f_i is the fitted value for age i.

With the matrix structure W_{V1}, the transmission is the highest between hosts aged 12–18 years ($54{\cdot}5 \times 10^{-5}$, see also Fig. 2). This implies that the highest probability of infection is between an infective aged 12–18 years and a susceptible belonging to the same age group, presumably because of the high frequency of close contacts. In the other age groups, the transmission increases almost monotonically with age for hosts aged $\leqslant$18 years and decreases monotonically with age for hosts aged $>$19 years. The transmission coefficient among hosts aged 31–44 years, which is also the 'background' transmission coefficient between hosts belonging to different age groups is very low ($1{\cdot}5 \times 10^{-5}$). The two highest transmission coefficients with the matrix structure W_{V2} are among hosts aged 2–5 years ($12{\cdot}7 \times 10^{-5}$) and between hosts aged 6 months–1 year and all hosts up to 6 years of age ($11{\cdot}6 \times 10^{-5}$). The mixing pattern with matrix W_{V3} is quite similar to the pattern with W_{V2}. The two highest transmission coefficients with W_{V3} is among hosts aged 2–5 years (14×10^{-5}) and between hosts aged 6 months–1 year and all hosts up to 12 years of age ($10{\cdot}8 \times 10^{-5}$). With matrix structure W_{V4}, the two highest transmission coefficients are for susceptible hosts aged 2–5 years ($10{\cdot}7 \times 10^{-5}$) and hosts aged 6 months–1 year ($10{\cdot}2 \times 10^{-5}$). The transmission coefficient decreases monotonically with increasing age for susceptible hosts aged $>$2 years. Below 18 years of age, W_{V5} and W_{V1} have similar mixing patterns. Transmission is highest between hosts aged 12–18 years ($66{\cdot}7 \times 10^{-5}$) and increases almost monotonically with age for hosts aged $\leqslant$18 years and decreases monotonically with age for hosts aged $>$19 years. However, the transmission coefficient in the 19–30 years age group, which is constrained to be equal to the parameter in the 31–44 years group is much higher than with matrix W_{V1}. Just as with W_{V1}, the 'background' transmission coefficient between hosts belonging to different age groups is very low ($0{\cdot}9 \times 10^{-5}$). With matrix W_{V6} with no transmission at all between hosts belonging to

different age groups, the transmission coefficient increases almost monotonically with age and is the highest in hosts aged 31–44 years ($162{\cdot}4 \times 10^{-5}$).

Estimation of R_0 from the WAIFW matrices

The global basic reproductive number R_0 for the population is the dominant eigenvalue of the 'next generation matrix' whose elements are the individual basic reproductive numbers $R_{0_{ij}}$ ($i = 1, 2, \ldots, 6$, $j = 1, 2, \ldots, 6$) for the transmission of the infection from an infectious person in the age group j to a susceptible person in the age group i. By definition of the basic reproductive number, each $R_{0_{ij}} = \beta_{ij} \times N_i \times D$, where β_{ij} is the ijth element of the WAIFW matrix, D is the mean duration of infectiousness and N_i is the total population in age group i. For varicella, the mean duration of infectiousness is 7 days = 7/365 years. Like the elements of the WAIFW matrix, the 95% CIs for R_0 are computed using two different methods: percentile non-parametric bootstrap and percentile parametric bootstrap.

The minimal immunization coverage needed for elimination, i.e. the proportion of the total population to be immunized immediately after waning of maternal antibodies in order to eliminate varicella, p_c, is obtained by the relationship:

$$p_c = 1 - \frac{1}{R_0}. \tag{16}$$

Bootstrap confidence interval for estimation of R_0 and p_c

Table 2 and Figure 3 show the parameter estimates for R_0 and p_c. Depending on the configuration of the WAIFW matrix, the basic reproductive number ranges between 3·12 (95% non-parametric CI 2·78–3·50) for W_{V3} to 68·57 (95% non-parametric CI 43·64–111·20) for the assortative mixing pattern W_{V6}. This implies that across the different configurations of the WAIFW matrices p_c ranges from 63·99% (lower limit for W_{V3}) to 99·10% (upper limit for the assortative mixing pattern).

Table 2. *Estimates of the basic reproductive number R_0 and p_c for different matrix structures for the 'Who Acquires Infection From Whom' (WAIFW) matrix*

			95% confidence intervals	
Parameter	WAIFW	Estimate	Non-parametric	Parametric
R_0	W_{V1}	11·99	8·29–17·98	8·75–18·34
	W_{V2}	3·19	2·87–3·57	2·89–3·54
	W_{V3}	3·12	2·78–3·50	2·79–3·49
	W_{V4}	4·34	3·90–4·89	3·90–4·81
	W_{V5}	26·33	18·95–37·79	12·29–38·17
	W_{V6}	68·57	43·64–111·20	43·46–118·64
p_c (%)	W_{V1}	91·66	87·94–94·44	88·57–94·55
	W_{V2}	68·65	65·12–71·99	65·35–71·75
	W_{V3}	67·92	63·99–71·40	64·11–71·31
	W_{V4}	76·95	74·39–79·55	74·38–79·19
	W_{V5}	96·20	94·72–97·35	94·82–97·38
	W_{V6}	98·54	97·71–99·10	97·70–99·16

DISCUSSION

When the force of infection is both time and age dependent, the WAIFW matrix is a central parameter in modelling the spread of the infection in the population. A structure has to be assumed for the WAIFW matrix in order to be able to estimate the transmission coefficients and different structures lead to different estimates for the basic reproductive number R_0 and the minimal immunization coverage needed for elimination of the infection in the population p_c. In this paper, we have estimated R_0 and p_c for varicella in Belgium for different configurations of the WAIFW matrix. First, the force of infection has been estimated from seroprevalence data stratified by age, using a parametric model with fractional polynomials. The estimates of the mean force of infection over six age groups has then given the means to estimate the transmission coefficients, R_0 and p_c for six different configurations of the WAIFW matrix. The variability of these parameters has been estimated through the computation of bootstrap confidence intervals. The results show that the values of R_0 and p_c are sensitive to the structure of the WAIFW matrix, with the estimates of R_0 ranging between 3·12 and 68·57 and those of p_c ranging between 67·92% and 98·54% for the six configurations chosen. Preliminary empirical data gathered by surveys about the mixing patterns for directly transmitted infections like varicella tend to show that people mostly mix with other people of the same age (configurations W_{V1}, W_{V5}, and W_{V6}). However, although it has the advantage of providing an upper bound on the values of R_0 and p_c, configuration W_{V6} is not realistic since people do not mix exclusively with people of the same age group. Moreover, the assumption of W_{V4}, that only the age of the susceptible hosts matters and not the age of the infectious hosts seems *a priori* unrealistic.

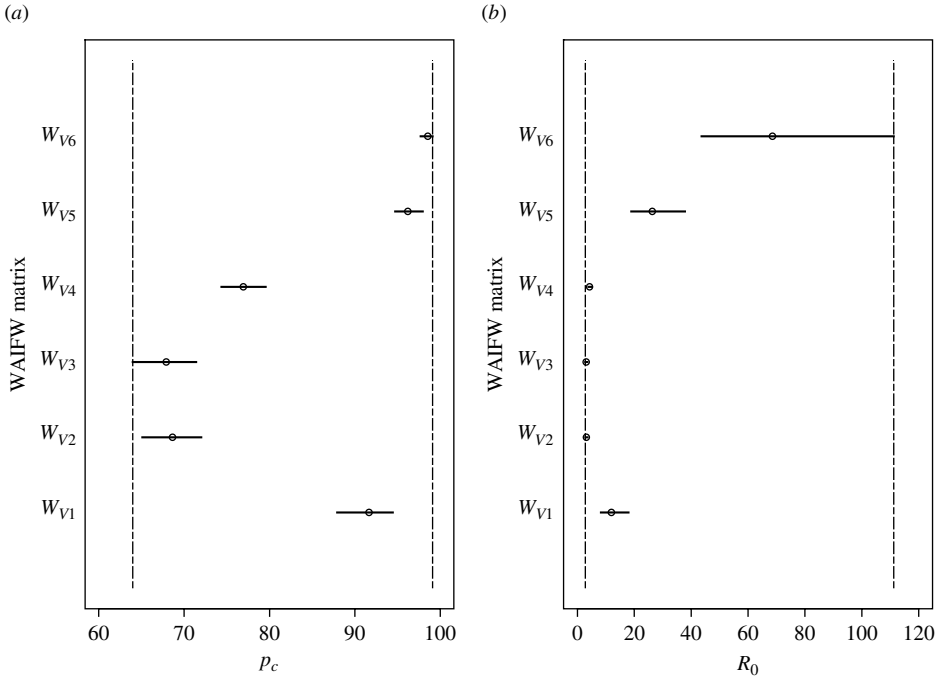

Fig. 3. Estimates for (*a*) p_c and (*b*) R_0 and non-parametric bootstrap confidence intervals.

On the other hand, mixing patterns like W_{V2} and W_{V3} are probably realistic for a childhood disease like varicella for which it can reasonably be assumed that transmission takes place mainly amongst groups of young children. Hence, configurations W_{V1}, W_{V2}, W_{V3}, and W_{V5} are probably the most relevant for varicella, which tends to support a value of R_0 between 3·12 and 26·33 and a value of p_c between 67·92% and 96·20%. Although using a different model with a force of infection that varies over time and only five age groups, Whitaker & Farrington [16] obtained similar values for R_0 for varicella in the United Kingdom. The estimates they obtained with a WAIFW matrix with a configuration similar to our W_{V2} and W_{V3} matrices were 3·02 and 3·14 in 1970 and 1998, respectively, while their estimate was 11·79 at both time-points for a WAIFW configuration similar to our assortative W_{V1} matrix. For elimination to be possible at a critical uptake, u_c, the vaccine's lifelong efficacy, e, should be between 64·0% and 97·4%, respectively, in order for elimination of varicella by vaccination to be theoretically possible (as $u_c = p_c/e$). Different means can be investigated to determine more precisely the value of R_0 and p_c. Gathering empirical data about mixing patterns [17, 18] should help us to better determine which is the most plausible configuration for a given infection in a population. Another possible avenue is to estimate R_0 and p_c using seroprevalence data from different infections that have a similar type of transmission, e.g. varicella, parvovirus, measles, mumps and rubella. Assuming a symmetric WAIFW matrix, seroprevalence data from three or four different infections would be enough to estimate the 15 (or 21), transmission coefficients with five (or six), age groups without additional assumptions about the mixing pattern. In any case, it seems clear that further studies on the most appropriate configuration of the WAIFW matrix are necessary to reduce variation in estimated R_0 and associated parameters.

APPENDIX

Expression of the β_is for the structure W_{V3}:

$$\beta_6 = \frac{\frac{L}{DN}\hat{\lambda}_6}{\Psi_1+\Psi_2+\Psi_3+\Psi_4+\Psi_5+\Psi_6}$$

$$\beta_5 = \frac{\frac{L}{DN}\hat{\lambda}_5-\beta_6\Psi_6}{\Psi_1+\Psi_2+\Psi_3+\Psi_4+\Psi_5}$$

$$\beta_4 = \frac{\frac{L}{DN}\hat{\lambda}_4-\beta_5\Psi_5-\beta_6\Psi_6}{\Psi_1+\Psi_2+\Psi_3+\Psi_4}$$

$$\beta_1 = \frac{\frac{L}{DN}\hat{\lambda}_1-\beta_4\Psi_4-\beta_5\Psi_5-\beta_6\Psi_6}{\Psi_1+\Psi_2+\Psi_3}$$

$$\beta_3 = \frac{\frac{L}{DN}\hat{\lambda}_3-\beta_1\Psi_1-\beta_4\Psi_4-\beta_5\Psi_5-\beta_6\Psi_6}{\Psi_2+\Psi_3}$$

$$\beta_2 = \frac{\frac{L}{DN}\hat{\lambda}_2-\beta_1\Psi_1-\beta_3\Psi_3-\beta_4\Psi_4-\beta_5\Psi_5-\beta_6\Psi_6}{\Psi_2}$$

ACKNOWLEDGEMENTS

This paper is based on T.v.E.'s M.Sc. thesis in Biostatistics, at Hasselt University. Z.S., M.A., P.V.D. and P.B. acknowledge funding from 'SIMID', a strategic basic research project of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), project no. 060081 and POLYMOD, a European Commission project funded within the Sixth Framework Programme, Contract no.: SSP22-CT-2004-502084. Z.S., M.A. and G.M. gratefully acknowledge the financial support from the IAP research network no. P5/24 of the Belgian Government (Belgian Science Policy).

DECLARATION OF INTEREST

This work is unrelated to T.v.E.'s work at GlaxoSmithKline Biologicals where he is employed as a mathematical modeller.

REFERENCES

1. **Anderson RM, May RM.** *Infectious Diseases of Humans, Dynamic and Control*. Oxford University Press, 1991.
2. **Halloran ME, Watelet L, Struchiner CJ.** Epidemiologic effects of vaccinations with complex direct effects in age-structured population. *Mathematical Biosciences* 1994; **121**: 193–225.
3. **Anderson RM, May RM.** Age-related changes in the rate of disease transmission: implication for the design of vaccination programmes. *Journal of Hygiene* (*Cambridge*) 1985; **94**: 365–436.
4. **Farrington CP, Kanaan MN, Gay NJ.** Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data (with Discussion). *Journal of the Royal Statistical Society, Series C: Applied Statistics* 2001; **50**: 251–292.
5. **Greenhalgh D, Dietz K.** Some bounds on estimation for reproductive ratios derived from the age-specific force of infection. *Mathematical Biosciences* 1994; **124**: 9–57.
6. **Wallinga J, *et al.*** Estimation for measles reproduction ratios and prospects for elimination of measles by vaccination in some western european countries. *Epidemiology and Infection* 2001; **127**: 281–295.
7. **Greenhalgh, D.** Discussion in Farrington *et al.* (2001), Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Journal of the Royal Statistical Society, Series C: Applied Statistics* 2001; **50**: 251–292.
8. **Keiding N.** Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society, Series A* 1991; **154**: 371–412.
9. **Shkedy Z, *et al.*** Modeling hepatitis A force of infection using monotone local polynomials. *Journal of the Royal Statistical Society, Series C: Applied Satistics* 2003; **52**: 469–485.
10. **Grenfell BT, Anderson RM.** The estimation of age-related rates of infection from case notification and serological data. *Journal of Hygiene* (*Cambridge*) 1985; **95**: 419–436.
11. **Farrington CP.** Modeling forces of infection for measles, mumps and rubella. *Statistics in Medicine* 1990; **9**: 953–967.
12. **Keiding N, *et al.*** Estimation from current status data in continuous time. *Lifetime Data Analysis* 1996; **2**: 119–129.
13. **Shkedy Z, *et al.*** Modeling age dependent force of infection from prevalence data using fractional polynomials. *Statistics in Medicine* 2006; **25**: 1577–1591.
14. **Thiry N, *et al.*** The seroepidemiology of primary varicella-zoster virus infection in Flanders (Belgium). *European Journal of Pediatrics* 2002; **161**: 588–593.
15. **Royston P, Altman DG.** Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling. *Applied Statistics* 1994; **43**: 429–467.
16. **Whitaker HJ, Farrington CP.** Infections with varying contact rates: application to varicella. *Biometrics* 2004; **60**: 615–623.
17. **Edmunds WJ, O'Callahgan CJ, Nokes DJ.** Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society of London, Series B* 1997; **264**: 949–957.
18. **Beutels P, *et al.*** Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiology and Infection* 2006; **134**: 1158–1166.