*Epidemiol. Infect.* (2004), **132**, 433–441. © 2004 Cambridge University Press
DOI: 10.1017/S0950268804001980 Printed in the United Kingdom

# Estimating the transmission parameters of pneumococcal carriage in households

A. MELEGARO[1,2]*, N. J. GAY[1] AND G. F. MEDLEY[2]

[1] *Modelling and Economics Unit, Communicable Disease Surveillance Centre, Health Protection Agency, London NW9 5EQ, UK*
[2] *Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK*

(*Accepted 22 December 2003*)

## SUMMARY

This paper analyses *Streptococcus pneumoniae* transmission dynamics in households using longitudinal data on pneumococcal (Pnc) carriage in the United Kingdom. Ten consecutive swabs were taken at 4-week intervals from all members of 121 households. The family status is derived from the observed Pnc carriage status of each family member. Transition matrices are built for each family size and composition containing the observed frequency of transitions between family statuses over a 28-day interval. A density-dependent transmission model is fitted to derive maximum-likelihood estimates of the duration of carriage and acquisition rates from the community and from infected individuals within the household. Parameter values are estimated for children (<5 years) and adults (5+ years). The duration of carriage is longer in children <5 years of age than in older family members (51 *vs.* 19 days). Children are 3–4 times more likely than adults to acquire Pnc infection from the community. Transmission rates within the household suggest that adults are more infectious but less susceptible than children. Transmission within the household is most important in large families. The proportion of household-acquired infection ranges from 29 to 46% in households of three persons to 38–50% in larger households. Evidence of density-dependent within-household transmission is found, although the strength of this relationship is not clear from the model estimates.

## INTRODUCTION

### Background

*Streptococcus pneumoniae* is one of the most important bacterial pathogens in respiratory tract infections, affecting children and adults worldwide despite current antibiotic therapies [1, 2]. It is responsible for causing upper respiratory infections which may lead to ear infections and sinusitis, as well as more serious invasive disease such as pneumonia, septicaemia and meningitis [3, 4]. Pneumococcus is the leading cause of lobar pneumonia in children under 5 years of age [5].

The bacterium gains entry into the host by colonizing the nasopharynx and the outcome of colonization depends on both the virulence of the colonizing serotype and the efficiency of the host immune system. The interval between colonization and onset of disease, when it occurs, is variable, but there is some evidence that disease is more likely to occur shortly after colonization [6]. However, most of pneumococcal (Pnc) infections remain asymptomatic and *S. pneumoniae* is considered a common component of the nasopharyngeal flora in healthy individuals.

* Author for correspondence: A. Melegaro, Modelling and Economics Unit, Communicable Disease Surveillance Centre, Health Protection Agency, 61 Colindale Avenue, London NW9 5EQ, UK.

Transmission from a person infected with pneumococcus is via droplets of respiratory secretions that remain air-borne over a distance of a few feet. The infecting organism may also be carried on hands contaminated with secretions. As the majority of Pnc-infected individuals remain asymptomatic, carriage data are needed in order to gain an in-depth understanding of Pnc transmission dynamics.

A heptavalent protein–polysaccharide conjugate vaccine has been licensed in the United Kingdom and it is now recommended for children at high risk under the age of 2 years. The vaccine has been proven to be highly effective against invasive Pnc disease [7] and moderately effective against non-invasive Pnc conditions (otitis media, pneumonia) [8–10]. Moreover, some efficacy against Pnc carriage has been shown [11], although there are still concerns about whether serotype replacement will occur if vaccination is widely introduced and what effects this will have on the severity of the disease. Baseline information on the transmission dynamics of Pnc carriage in the pre-vaccination era is thus necessary to predict the effects of interventions such as the introduction of the vaccine on disease burden.

### Previous work

Longitudinal carriage studies have been performed to gain insight into the pathogen's mechanism of carriage and transmission within hosts. In this type of study, individuals are swabbed at regular intervals, their carriage status is assessed and inference on their rate of acquisition and on the duration of carriage can be obtained. Smith et al. [12] using longitudinal carriage data collected in Papua New Guinea found that Pnc serotypes vary widely in their acquisition rates and in the length of time for which they persist. An inverse relationship between duration of carriage and age was shown by Ekdahl et al. [13] who studied Pnc carriage using data from an intervention project in Sweden. A similar relationship was found by Auranen et al. [14] who analysed longitudinal data on Pnc carriage in Finnish families taking into consideration transmission within the family and from the community and estimating acquisition rates and duration of carriage for children and adults.

### The problem

The aim of this work is to analyse data from a longitudinal carriage study in families in the United Kingdom thereby gaining insights into Pnc transmission dynamics in the pre-vaccination era. In particular, to estimate the duration of carriage for adults and children, and the rate of acquisition of carriage from the community and from infected family members.

## METHODS

### Description of the data

The dynamics of Pnc transmission within families were captured by a longitudinal study of Pnc carriage conducted in the United Kingdom from October 2001 to July 2002 as part of the European Pneumococcal Project (PncEuro) [15]. The data consisted of follow-up measurements of Pnc carriage of 132 pre-school children (<3 years) and their families, who were enrolled through primary health-care registers in Hertfordshire [15]. The size of the family varied from 2 to 7, although in most of the families there were 3 or 4 members.

From each family member, a nasopharyngeal swab was obtained at initial home visit and followed by 9 further swabs at 4-week intervals. As a previous study [13] suggested that duration of carriage might be less than 28 days, especially in adults, it is possible that not all episodes of carriage will be observed from the data. Although serotyping was performed when Pnc carriage was detected, this analysis considers Pnc carriage in general and, thus, is only looking at whether or not the individual is carrying the bacteria. As there were only three circumstances in which two serotypes were found in the same individual at the same time, the model is based on the assumption that simultaneous carriage of different serotypes is rare enough to ignore initially.

A total of 121 families comprising 489 individuals were considered in this analysis. 11 families dropped out from the study at the very beginning. For each individual that took part in the study, the carriage status at each monthly visit was obtained from the data and then recoded as 0, if non-carrier, 1 if carrier and 9 when either the swab was not taken or the laboratory result not reported. The household state at each visit is derived combining the carriage status of all the family members and results in a sequence of 0s and 1s if their information is complete. Missing information on at least one family member results in incomplete information at the household level

https://doi.org/10.1017/S0950268804001980 Published online by Cambridge University Press

and, thus, a missing value code. Family sequences are organized so that the last number represents the status of the youngest member of the family. Due to the strong association that has been shown in other studies [13, 14] between Pnc carriage and age, with younger children having a much higher prevalence and longer duration of carriage, we set a cut-off at 5 years of age and stratified households by family size and composition (number of adults and children in the family). We derived strings of household results, which were then converted into a table showing the number of transitions between each pair of household states over a 28-day period for each family size and structure. Two households of 2 individuals and one of 7 were excluded from the analysis as their sequences were non-informative. Ten distinct tables $N_{z,u}$ were derived for the available combinations of household size ($z$) and number of adults ($u$) respectively in the family: (3, 1), (3, 2), (4, 2), (4, 3), (5, 2), (5, 3), (5, 4), (6, 3), (6, 4), (6, 5). Only complete family transitions, where the infection state of all household members is known on two consecutive observations, are used in the following analysis.

## The model

Following the work by Auranen and colleagues on both Hib and Pnc infection [14, 16] the model considers transmission of Pnc within the household. We assumed that each individual can either be in a non-infected state, i.e. susceptible to infection (S), or they can already be infected (carriers) (C) and, as a consequence, able to spread the infection to uninfected individuals in their household. No Pnc-specific disease was observed so this possibility is not included. As in Auranen et al. [14] we set the transition from C→S to be dependent on a constant recovery rate, and the transition from S→C to be a function of the potential exposure to the infective agent both within the family (within-family acquisition rate) and outside the family (community acquisition rate).

The probabilities of transition from an infected to an uninfected state and vice versa in a short time interval $\delta t$ are then defined for an individual in the age class $i=1, 2$ [where 1 = child (< 5 years) and 2 = adult (5+ years)]:

$$\mathrm{P}_i(\mathrm{C}\rightarrow\mathrm{S})_{\delta t}=\mu_i\cdot\delta t, \tag{1}$$

$$\mathrm{P}_i(\mathrm{S}\rightarrow\mathrm{C})_{\delta t}=\left(k_i+\frac{\beta_{i1}\mathbf{I_1}(t)+\beta_{i2}\mathbf{I_2}(t)}{(z-1)^w}\right)\cdot\delta t, \tag{2}$$

where $\mu_i$ and $k_i$ are the clearance and the community acquisition rates respectively for age class $i$, and $z$ is the family size. $\mathbf{I_1}(t)$ and $\mathbf{I_2}(t)$ are the number of infected children and infected adults respectively in the family. $\beta_{ij}$ is the transmission rate from an infected to an uninfected individual and it reflects both the infectiousness of an individual in age class $j$ and the susceptibility of an individual in age class $i$. Thus the within-household probability of being infected in $\delta t$ is a function of the number and age class of the other family members that are carrying the bacteria at that particular time. Community acquisition rates and recovery rates are similarly defined in vectors: $(k_1, k_2)$ and $(\mu_1, \mu_2)$. The term $(z-1)^w$ in eqn (2) represents a density correction factor; $(z-1)$ being the number of other family members in families of size $z$, and $w$ the factor that controls the extent of density dependence. When $w=0$ the model reflects density-independent transmission: the probability of contracting infection from another specific individual in the family does not change with family size and, thus, the average number of contacts increases linearly with family size. When $w=1$ the probability of getting in contact with the same individual decreases in bigger families although the average number of contacts remain constant. If $w>1$ the average number of contacts decreases with family size. In total nine parameters define the model.

Equations (1) and (2) are used to derive the probability $P_{rs}$ of transition between two different family states $r$ and $s$ in a short time-interval $\delta t$. These probabilities are calculated for each household size and composition. It is assumed that in a short time-interval $\delta t$, here chosen as 1 day, only one member of the household will change status (either through infection or recovery). The probability of transitions where more than one individual changes status is therefore set to zero. For example, in a family size of three composed of two adults and one child, the probabilities of going from state 001 (two susceptible adults and one carrier child) to states 000, 101 and 011 in one time-step, are derived using respectively eqn (1) to become a fully susceptible household, and eqn (2) to reach states 101 or 011. The $\mathbf{T_1}(3, 2)$ matrix [see eqn (3) below], contains all the probabilities of transitions from the initial family status at the top of the matrix to the consecutive status given on the left-hand side

https://doi.org/10.1017/S0950268804001980 Published online by Cambridge University Press

of the matrix when $\delta t$ is 1 day:

$$\begin{array}{c} \\ 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{array} \begin{array}{c} \begin{array}{cccccccc} 000 & 100 & 010 & 110 & 001 & 101 & 011 & 111 \end{array} \\ \begin{bmatrix} q_{11} & \mu_2 & \mu_2 & 0 & \mu_1 & 0 & 0 & 0 \\ k_2 & q_{22} & 0 & \mu_2 & 0 & \mu_1 & 0 & 0 \\ k_2 & 0 & q_{33} & \mu_2 & 0 & 0 & \mu_1 & 0 \\ 0 & [2^{-W}\beta_{22}+k_2] & [2^{-W}\beta_{22}+k_2] & q_{44} & 0 & 0 & 0 & \mu_1 \\ k_1 & 0 & 0 & 0 & q_{55} & \mu_2 & \mu_2 & 0 \\ 0 & [2^{-W}\beta_{12}+k_1] & 0 & 0 & [2^{-W}\beta_{21}+k_2] & q_{66} & 0 & \mu_2 \\ 0 & 0 & [2^{-W}\beta_{12}+k_1] & 0 & [2^{-W}\beta_{21}+k_2] & 0 & q_{77} & \mu_2 \\ 0 & 0 & 0 & [2^{1-W}\beta_{12}+k_1] & 0 & [2^{-W}(\beta_{22}+\beta_{21})+k_2] & [2^{-W}(\beta_{22}+\beta_{21})+k_2] & q_{88} \end{bmatrix} \end{array} \quad (3)$$

The sensitivity of the model to changes in the definition of $\delta t$ ($\frac{1}{2}$ day, 2 days) was checked. The daily probability that the family does not change state is given by $q_{rr}=1-\Sigma_{s\neq r}P_{rs}$. Similar formulations of the transition matrix $\mathbf{T}_1$ are expressed for families with 3–6 members and different age structures.

The families transition probabilities for a 28-day interval were then derived as $\mathbf{T}_{28}(3, 2)=\mathbf{T}_1^{28}(3, 2)$. This allowed handling transitions in which more than one individual in the family changes status. Moreover, not having specified the pathway from one state to the next, unobserved events were implicitly included.

## Parameter estimation

Maximum-likelihood techniques are adopted to estimate the nine model parameters. The log likelihood $L(z, u)$ and saturated log likelihood $L^*(z, u)$ functions are derived from each monthly transition matrix $\mathbf{T}_{28}(z, u)$ and from the observed number of family transitions between states $n_{rs}(z, u)$. The deviance is then calculated for each family size and composition as follows:

$$\mathrm{Dev}_{z,u}=$$

$$2\cdot(L^*-L)=2\cdot\sum_{r=1}^{2^z}\sum_{s=1}^{2^z}n_{rs}[\ln(n_{rs})-\ln(P_{rs}N_r)], \quad (4)$$

where $n_{rs}$ is the number of family transitions observed in the data from state $r$ to state $s$ and $N_r=\Sigma n_{rs}$. The $P_{rs}$ is the element $rs$ of matrix $\mathbf{T}_{28}(z, u)$ or, in other words, is the model probability that a family of size $z$ and with $u$ adults moves from state $r$ to state $s$ in a 28-day interval. The upper limit of the summations represent the number of family states that are possible for each family size.

The overall deviance [eqn (5)] is obtained as the sum over all family sizes and compositions

$$\mathrm{Dev}=\sum_{z=3}^{6}\sum_{u=1}^{z}\mathrm{Dev}_{z,u}. \quad (5)$$

Maximum-likelihood estimates for the parameters are obtained by minimizing the deviance.

The profile-likelihood method [17] is used to derive confidence intervals for the model parameters. The analysis was conducted using Microsoft Excel.

## Model fit

The overall fit of the model could not be formally assessed using a $\chi^2$ test on the residual deviance and degrees of freedom. This was because the expected count for each cell was too small and usually less than 1.

However, after aggregating the states with the same number of carriers, it was possible to assess the goodness of fit visually for each family size by plotting the number of family transitions between states and comparing it to the number obtained from the model after aggregating across cells. $\chi^2$ tests comparing the observed and estimated number of transitions on the aggregated data were performed. Using the $\chi^2$ test in this way only enables the detection of fairly large aberrations in the model fit because the fit is only being tested after the aggregation of data and estimates from a more complex model.

## Prevalence of Pnc infection by household size and composition

A steady-state vector $\mathbf{v}$ of dimension $2^z$ is calculated from any of the family transition matrix [i.e. $\mathbf{T}_1(z, u)$]

https://doi.org/10.1017/S0950268804001980 Published online by Cambridge University Press

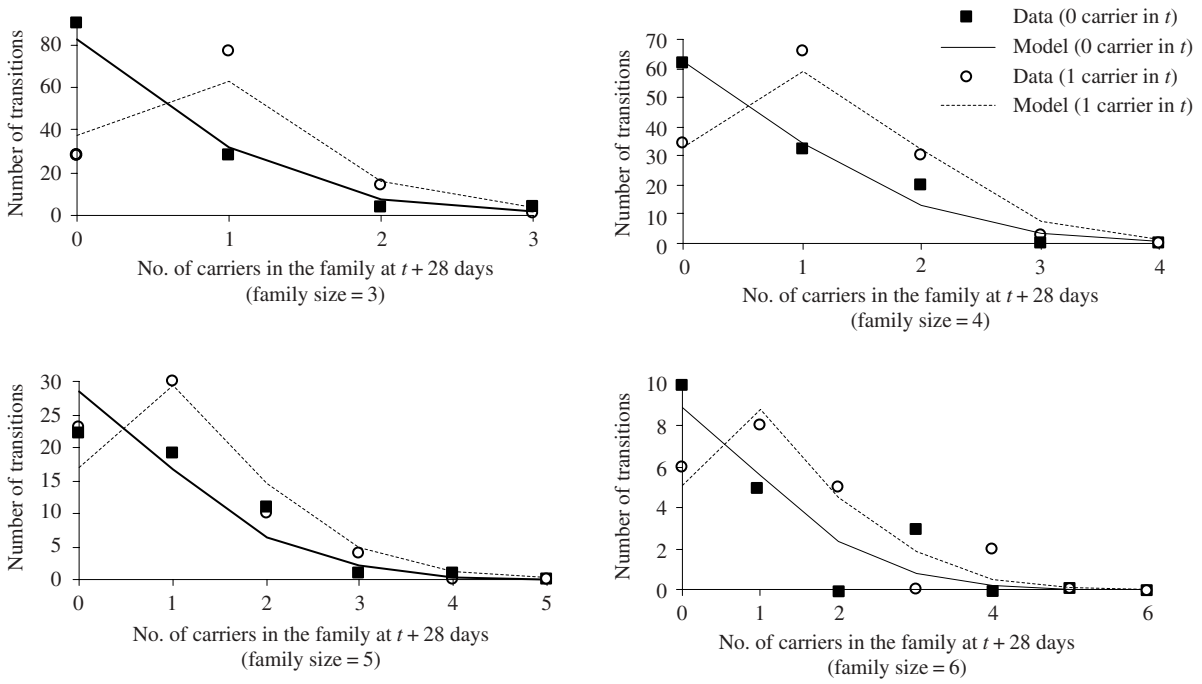


**Fig. 1.** Transition probabilities obtained from the model for completely susceptible families (——) and families with one carrier (- - - -), compared to proportions of transition observed from carriage data (○, ■).

by solving the following system of equations

$$\mathbf{T_1}(z,u)\cdot\mathbf{v}(z,u)=\mathbf{v}(z,u), \tag{6}$$

where **v** gives the proportion of households that are in each specific state at the equilibrium [18]. From this the expected prevalence of infection in children and adults in a household of given size and structure can be calculated.

The proportion of community-acquired Pnc infection is calculated using the steady-state vector **v** and deriving the expected number of new infections produced exclusively by household transmission ($k=0$) and comparing it to the expected number of those acquired from the community ($\beta=0$).

## RESULTS

We had 560 (51%) complete family transitions in the data (Table 1), the majority of which are found in a family size of 3 and 4. The overall deviance obtained comparing the model log likelihood to the saturated log likelihood is 645, with 180 D.F. (189 non-zero datapoints and nine parameters to estimate). In Figure 1 the fit of the model is shown comparing the expected number of transitions estimated by the model, for uninfected families and families with one carrier, to the number observed from the data for the same transitions. The $\chi^2$ tests of the model fit (not shown) did not show evidence of any significant difference between the model and the data.

Table 1. *Number of complete transitions by family size and number of adults in the family*

| Family size | No. of adults | | | | | Total |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| 3 | 8 | 226 | — | — | — | 234 |
| 4 | — | 146 | 54 | — | — | 200 |
| 5 | — | 8 | 27 | 45 | | 80 |
| 6 | — | — | 14 | 1 | 31 | 46 |
| Total | 8 | 380 | 95 | 46 | 31 | 560 |

Estimates of the nine model parameters are shown in Table 2. In children <5 years the mean duration of Pnc carriage, derived using the inverse of the estimated recovery rate, is 51 days (95% CI 42–64). The estimated mean duration of carriage in older family members is close to 19 days (95% CI 14–24). The community acquisition rate for children (0·012 per day) is more than 3 times higher than that for adults (0·004 per day), showing that children <5 years are most likely to introduce the infection into the household. On the other hand, estimates of the within-family acquisition rates show that the highest

https://doi.org/10.1017/S0950268804001980 Published online by Cambridge University Press

Table 2. *Maximum-likelihood parameter estimates*

| Description of model parameters | Symbol used | Rates (per day) | 95% CI |
|---|---|---|---|
| Community acquisition rate | | | |
| Adult | $k_2$ | 0·004 | 0·002–0·005 |
| Child | $k_1$ | 0·012 | 0·008–0·016 |
| Within-family acquisition rate | | | |
| Adult to adult | $\beta_{22}$ | 0·048 | 0·010–0·180 |
| Adult to child | $\beta_{12}$ | 0·106 | 0·020–0·450 |
| Child to adult | $\beta_{21}$ | 0·005 | 0·000–0·018 |
| Child to child | $\beta_{11}$ | 0·047 | 0·008–0·200 |
| Recovery rate | | | |
| Adult | $\mu_2$ | 0·053 | 0·041–0·070 |
| Child | $\mu_1$ | 0·020 | 0·016–0·024 |
| Density factor | $w$ | 1·184 | 0·200–2·200 |

CI, Confidence interval.

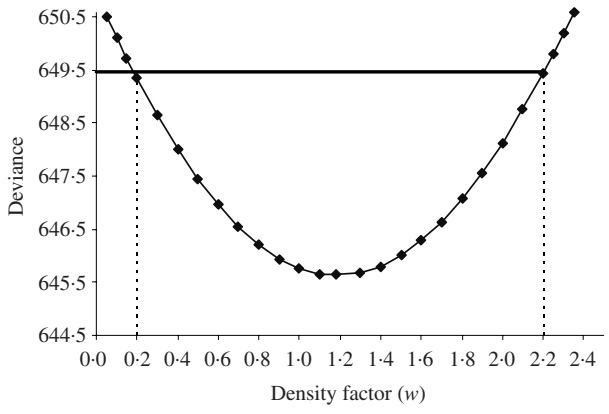


**Fig. 2.** Profile likelihood for the density factor $w$ (the dotted lines represent the 95% confidence interval).

daily transmission rate is that from adults to children, whereas the lowest is the one from children to adults. These results are consistent with adults being less susceptible to infection. A density coefficient significantly greater than zero ($w = 1{\cdot}2$, 95% CI 0·2–2·2) is estimated from the model and suggests the importance of considering density-dependent transmission. However, due to the very wide confidence interval, no clear indication is given on the strength of this effect (Fig. 2). The sensitivity of the model to variation in the time-interval $\delta t$ ($\delta t = \frac{1}{2}$ day, $\delta t = 2$ days) was assessed and the parameters estimates were found to be not significantly different from the base case results ($\delta t = 1$ day).

The expected prevalence of Pnc infection at equilibrium in both adults and children is given in Table 3.

Table 3. *Expected equilibrium prevalence of pneumococcal carriage in household with different compositions (children 0–5 years; adults 5+ years) ($w = 1$)*

| Prevalence | No. of children in household | No. of adults in household 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Adults | 0 | 0·06 | 0·10 | 0·11 | 0·12 | 0·12 |
| | 1 | 0·09 | 0·10 | 0·11 | 0·11 | 0·12 |
| | 2 | 0·10 | 0·11 | 0·11 | 0·11 | — |
| | 3 | 0·10 | 0·11 | 0·11 | — | — |
| Children | 1 | 0·44 | 0·47 | 0·48 | 0·49 | 0·50 |
| | 2 | 0·54 | 0·53 | 0·52 | 0·52 | — |
| | 3 | 0·58 | 0·56 | 0·55 | — | — |

Table 4. *Estimated proportion of pneumococcal infection in adults and children acquired within the family for different household size and composition ($w = 1$)*

| | No. of children in household | No. of adults in household 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Adult | 0 | 0% | 40% | 47% | 50% | 52% |
| | 1 | 34% | 41% | 45% | 48% | 50% |
| | 2 | 40% | 42% | 45% | 47% | — |
| | 3 | 42% | 43% | 45% | — | — |
| Child | 1 | 22% | 29% | 33% | 36% | 38% |
| | 2 | 46% | 43% | 43% | 43% | — |
| | 3 | 54% | 50% | 49% | — | — |

In order to have a baseline probability, the expected Pnc carriage in adults is estimated also for families with no children although no such families exist in our data. The prevalence of Pnc carriage in adults does not vary much when bigger families are considered or when more children are present in the family. The expected prevalence in children increases with the number of children in the household. The prevalence increases by 3–10% for each additional child.

The proportion of household-acquired Pnc infection increases in bigger families, reaching 50% for both children and adults (Table 4). When only one child is in the family the proportion in children (22–38%) is lower than in adults (34–50%).

## DISCUSSION

Data from a UK Pnc longitudinal carriage study were used and a transmission model fitted using maximum-likelihood techniques to gain some insights into Pnc

https://doi.org/10.1017/S0950268804001980 Published online by Cambridge University Press

transmission dynamics within the household, shedding some light on its mechanism of acquisition and on the duration of carriage. We used these estimates to derive the individual equilibrium prevalence of Pnc infection in households of given size and structure and to assess the importance of household *vs*. community acquisitions.

Although the study was carefully designed and the swabbing interval was forced to its minimum, the presence of unobserved events had to be considered in the setting up of the model. In the past, Bayesian data augmentation methods have been adopted in order to tackle this issue and Markov Chain Monte Carlo (MCMC) simulation has been used to explore the joint posterior of the model parameters and the augmented data [14]. Here we decided to deal with unobserved events by setting up the model for a short time-interval (1 day), during which we assume only one event occurs, and then deriving the probabilities of transitions for longer (i.e. 28 days) intervals. This enabled us to describe the entire data-set with only nine model parameters.

The model classifies individuals as Pnc carrier or non-carrier and it does not consider which serotype they were carrying. Acquisition and clearance rates are thus estimated for carriage of any Pnc serotype. Serotype-specific differences in Pnc acquisition and clearance rates were shown in the past [12] and represent a very critical issue in terms of future disease burden and impact of vaccination policies. Here some baseline information on Pnc transmission dynamics are derived and, although the model allows successive positive tests to be attributed to recovery and reacquisition, it does not exploit the information provided by a change of the serotype carried. Further work will be pursued on these matters.

In agreement with previous findings, the model estimated that children carry Pnc for longer periods than adult family members (51 *vs*. 19 days). Smith et al. [12] estimated approximately 2 months of carriage in children for the most common serotypes (6, 19, 23) and Auranen et al. [14] published an estimate of 2·3 months (95% CI 1·5–3·3), derived for children aged $<2$ years. The difference between the latter estimate and ours may be due to the younger age group considered. Similarly, the estimated 1·5 months (95% CI 1·0–2·0) duration of carriage derived from the Finnish study [14] in children aged 2+ years is higher than our estimate in the 5+ years group, and this, again, is probably due to their being a younger age group. Moreover, their study allowed for control of exposures but the estimates produced may be imprecise as the study used long intervals between observations (1 month until 6 months of age, and then 3 and 6 months). Slightly lower estimates of the duration of carriage were found in the study of Ekdahl et al. [13], where a clear relationship between carriage and age was observed in individuals with clinical diagnosis of pneumonia and otitis media who were carrying penicillin-resistant Pnc strains. A median of 30 days for children aged $<1$ year old; 21 days for children 1–2 years; 21 days for children 3–4 years; 12 days for children 5–6 years; 15 days for persons 7–18 years and 14 days for adults 18+ years.

Estimates of Pnc acquisition rates from the community show a difference between children and adults. The former appear 3–4 times more likely to acquire Pnc infection from the community and to introduce the bacteria in their household's environment. Previous work by Auranen et al. [14] showed a similar pattern although the difference between children and adults appears to be stronger in the Finnish data and their rates are significantly lower than ours. Although their model considers only three serotypes that account for 30–60% of infections, this does not fully explain the lower community acquisition rates that they obtain.

The within-family transmission rates incorporate infectiousness of the carrier, susceptibility of the non-carrier and contact between them. Although various individual factors, such as age and immunity levels, may contribute to this relationship, the model used here is from the SIS class, and thus, it does not explicitly account for immunity after infection. Nevertheless, transmission rates within the family were derived considering the age class to which both the carrier and the non-carrier belong ($<5$ years, 5+ years) and suggested a higher level of susceptibility in children and infectiousness in adults.

The importance of density dependence when looking at transmission dynamics of infection has been discussed previously [19, 20]. We included a density-dependent factor in our analysis and we found it to be an important element of the model and necessary to explain within-family transmission ($w$ significantly greater than 0). However, at this stage, no clear judgement can be made on whether the transmission increases ($w<1$) or decreases ($w>1$) as a consequence of being in bigger families. More work is clearly necessary to investigate this relationship and, thus, to provide better estimates of within-family transmission that takes into consideration changes in the contact

patterns in households of different sizes, environment and composition.

As expected, the prevalence of Pnc carriage for any family size and structure is higher for children than for adults. Moreover, we found that whereas the prevalence in adults does not vary much for different family sizes and composition, the equilibrium prevalence in children increases with the number of children and adults in the family. This may again reflect the lower level of susceptibility in older individuals and, thus, the fact that they are not sensitive to changes in the proportion of infectious contacts they may have within the household.

Household-acquired infection makes a major contribution to Pnc transmission in families both for children and adults. The more siblings that are present in the family the more the infection can circulate among them and persist within the household. However, even in large families approximately 50–60% of children and adults' infections are acquired outside the household.

Longitudinal data can be quite challenging to analyse and many critical aspects have to be taken into account in doing so. Nevertheless they can provide a great contribution to the understanding of a pathogen's dynamics within the human host and on the way individual properties can affect transmission dynamics of the infectious agent. In this paper we tried to draw some conclusions about Pnc transmission dynamics, the contribution of community *vs.* family transmission and the importance of family structure when considering prevalence of Pnc infection within the household. Further analysis is clearly needed in order to provide some specific understanding of serotype-specific Pnc carriage and to be able to give some baseline information on the links between serotypes and transmission in the pre-vaccination era.

## ACKNOWLEDGEMENTS

We thank Drs Elisabeth Miller and Robert George for providing us with pneumococcal carriage data and for giving helpful advice. We also thank Drs Richard Pebody and Mahein Hussain for organizing the study and for giving us continuous updates on its progress. Thanks are due to all the nurses for maintaining the study and assisting the families involved throughout the study. We are grateful to Drs John Edmunds, Martin Eichner, Kari Auranen and Nick Andrews for useful discussions on the manuscript. We thank the EU for financial assistance (QLG4-CT-2000-00640).

## REFERENCES

1. Fedson DS, Scott JA, Scott G. The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known. Vaccine 1999; **17**: S11–S18.
2. Robinson K, Baughman W, Rothrock G, et al. Epidemiology of invasive *Streptococcus pneumoniae* infections in the United States, 1995–1998. Opportunities for prevention in the conjugate vaccine era. JAMA 2001; **285**: 1729–1735.
3. Miller E, Waight P, Efstratiou A, Brisson M, Johnson A, George R. Epidemiology of invasive and other pneumococcal disease in children in England and Wales 1996–1998. Acta Paediatr Suppl 2000; **435**: 11–16.
4. Lloyd-Evans N, O'Dempsey TJ, Baldeh I, et al. Nasopharyngeal carriage of pneumococci in Gambian children and in their families. Pediatr Infect Dis J 1996; **15**: 866–871.
5. Djuretic T, Ryan M, Miller E, Fairley CK, Goldblatt D. Hospital admission in children due to pneumococcal pneumonia in England. J Infect 1998; **37**: 54–58.
6. Schaechter M, Engleberg N, Eisenstein BI, Medoff G. Mechanisms of microbial disease. 3rd edn. Baltimore: Lippincott Williams & Wilkins, 1998.
7. Black SB, Shinefield HR, Fireman B, et al. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Pediatr Infect Dis J 2000; **19**: 187–195.
8. Fireman B, Black SB, Shinefield HR, Lee J, Lewis E, Ray P. Impact of the pneumococcal conjugate vaccine on otitis media. Pediatr Infect Dis J 2003; **22**: 10–16.
9. Black SB, Shinefield HR, Ling S, et al. Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than five years of age for prevention of pneumonia. Pediatr Infect Dis J 2002; **21**: 810–815.
10. Eskola J, Kilpi T, Palmu A, et al. Efficacy of a pneumococcal conjugate vaccine against acute otitis media. N Engl J Med 2001; **344**: 403–409.
11. Dagan R, Givon-Lavi N, Zamir O, et al. Reduction of nasopharyngeal carriage of Streptococcus pneumoniae after administration of a 9-valent pneumococcal conjugate vaccine to toddlers attending day care centers. J Infect Dis 2002; **185**: 927–936.
12. Smith T, Lehmann D, Montgomery J, Gratten M, Riley ID, Alpers MP. Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. Epidemiol Infect 1993; **111**: 27–39.
13. Ekdahl K, Ahlinder I, Hansson HB, et al. Duration of nasopharyngeal carriage of penicillin-resistant

*Streptococcus Pneumoniae*: experiences from the South Swedish Pneumococcal Intervention Project. Clin Inf Dis 1997; **25**: 1113–1117.
14. Auranen K, Arjas E, Leino T, Takala AK. Transmission of pneumococcal carriage in families: a latent Markov Process Model for binary longitudinal data. J Am Stat Assoc 2000; **95**: 1044–1053.
15. Hussain M, Clark M, Tew A, Williams L, Miller E. Upper respiratory tract carriage and transmission of pneumococci in families in UK and possible introduction of the new pneumococcal conjugate vaccine. Nursing Times, in press.
16. Auranen K, Ranta J, Takala AK, Arjas E. A statistical model of transmission of Hib bacteria in a family. Stat Med 1996; **15**: 2235–2252.
17. Aitkin M. Profile likelihood. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: John Wiley and Sons, 1998: 3534–3536.
18. Chin Long Chiang. An introduction to stochastic processes and their applications. New York: Robert E. Krieger, 1980.
19. Finkenstadt B, Bjornstad O, Grenfell B. A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. Biostatistics 2002; **3**: 493–510.
20. McCallum H, Barlow N, Hone J. How should pathogen transmission be modelled? Trends Ecol Evol 2001; **16**: 295–300.