Pneumococcal Carriage in United Kingdom Families: Estimating Serotype-specific Transmission Parameters from Longitudinal Data

Alessia Melegaro\(^1\), Yoon Choi\(^1,2\), Richard Pebody\(^2\), and Nigel Gay\(^1\)

\(^1\) Modelling and Economics Unit, Centre for Infections, Health Protection Agency, London, United Kingdom.
\(^2\) Immunisation Department, Centre for Infections, Health Protection Agency, London, United Kingdom.

Revised for publication November 23, 2005; accepted for publication January 23, 2007.

Repeated observations of pneumococcal infection in 121 United Kingdom families (October 2001–July 2002) were used to explore the transmission properties of five highly prevalent pneumococcal serotypes (6A, 6B, 14, 19F, 23F). A family-based Markov model was developed, and maximum likelihood estimates were produced for model parameters. The authors found higher community acquisition rates among preschool children for all serotypes and higher within-household transmission for 6A and 14. Significant differences in the spontaneous clearance rate were estimated between age categories and serotypes, with 6B being carried for almost 4 months in children. Different mechanisms of competition between serotypes were investigated, and a complete exclusion model (i.e., the resident strain cannot be outcompeted by challengers) was discarded in favor of a competing mechanism that leaves a resident serotype partially or fully susceptible to challengers. Large variation was found in the challenging strength, which was low for 19F and 23F and high for 6A and 6B. Serotype 6B was the only one characterized by high resistance capacity. Only small differences in the transmission characteristics were found when vaccine and nonvaccine serotypes were grouped, suggesting that a serotype-specific analysis is needed to detect distinctive serotype behavior.

Streptococcus pneumoniae is one of the most important bacterial causes of respiratory tract infections, affecting children and adults worldwide (1–4). It is a common component of the nasopharyngeal flora in healthy individuals but is also responsible for causing a range of disease from upper respiratory tract infections, otitis media, and sinusitis to pneumonia, septicemia, and meningitis (5–8).

A heptavalent pneumococcal conjugate vaccine has been registered in the United States and recommended by the American Academy of Pediatrics and the Advisory Committee on Immunization Practices for universal use in children less than 24 months of age and also for high-risk children aged 2–5 years. It has also recently been introduced in a number of European countries, including the United Kingdom, for routine infant vaccination, with various schedules (4). However, a reduction in carriage of vaccine serotypes following introduction of vaccination may reduce pressure on nonvaccine types. Vaccine serotypes are replaced by nonvaccine serotypes in the nasopharyngeal niche of colonized individuals, causing a shift in the serotypes circulating in the population and thus in disease. Evidence of serotype replacement has already been observed in both carriage (9–13) and disease (13–19), although the real extent of this phenomenon and the overall effect on pneumococcal-related morbidity will not be clear for years after widespread introduction of the conjugate vaccine.

Mathematical modeling has been extensively used in the past to better understand the epidemiology of infectious diseases and the potential impact of public health interventions. However, to provide sound support to policy decision makers, these models should be based on realistic parameter estimates, ideally derived from data collected in the field.
where the intervention under consideration has taken place, and an as-thorough-as-possible understanding of the mechanisms through which the infection is spreading in the population. For pneumococcus, as for many other recurrent infections, the latter is difficult to achieve, and time-consuming and expensive longitudinal studies are needed to capture the dynamics of multiple acquisition and recovery events and to incorporate them in the analysis.

Baseline estimates of pneumococcal transmission parameters have been derived in previous work (20) using a United Kingdom–based longitudinal data set of pneumococcal carriage in households (21). With this analysis, we developed a modeling framework that enabled us to incorporate the two levels of clustering typical of panel family observations: repetitive information at the individual level and a clustering effect at the household level. However, although information on serotype was also available from the study, the analysis considered pneumococcal carriage in general, and the work focused primarily on development of the methodology (i.e., modeling household transmission) and on being able to detect age-related dependencies in the transmission and duration of carriage parameters.

In the current analysis, we aimed to expand the previous modeling framework to include serotype-specific information available for the longitudinal study in the United Kingdom. In particular, our objective was to ascertain whether there are differences in transmissibility between the most commonly carried pneumococcal serotypes (6A, 6B, 14, 19F, 23F) and to estimate serotype-specific duration of carriage in children (<5 years of age) and older individuals (≥5 years of age). Moreover, different mechanisms of competition between the serotypes were also considered in the current modeling framework, which will help in gaining more insights into existing serotype interactions and in furthering our knowledge on how serotype replacement might occur.

**MATERIALS AND METHODS**

**Data**

The data set consisted of measurements of pneumococcal carriage in 121 preschool children (<3 years of age) and all household members during a follow-up period of 10 months from October 2001 to July 2002 (21). Families were visited once per month, and nasopharyngeal swabs were taken from all family members, following the World Health Organization guidelines for detecting upper respiratory carriage of *S. pneumoniae* (22). Samples were cultured for pneumococcus according to standard methods and, if positive, were serotyped.

A nasopharyngeal swab was taken at least once in 489 individuals (121 families); a total of 3,753 swabs were taken, of which 932 (25 percent) revealed pneumococcal carriage. Thirty-four different serotypes were detected in the study population. Of these, the following five represented the majority of positive swabs (75 percent in children and 52 percent in adults): 6A, 6B, 14, 19F, and 23F. The maximum number of different serotypes cultured from an individual during the 10-month study period was five. One or two serotypes were detected in 222 individuals (45 percent), three serotypes in 35 individuals (7 percent), and four or five serotypes in 20 (4 percent). Children from 6 months to 4 years of age carried the maximum number of serotypes during the 10-month period. The most prevalent serotypes were detected six or more times in the same person (although not consecutively), but most types were found fewer than three times in the same person.

Consecutive episodes of carriage of the same serotype were frequent, especially for the most prevalent types and in children. This finding suggests that the duration of carriage may be longer at younger ages. However, to better ascertain the existence of differences among children and adults, modeling techniques were developed to account for the possibility of unobserved events that may have occurred in between swabbing intervals (i.e., loss and reacquisition) and that would have influenced the estimates.

**Preparation of data sets**

Five distinct data sets were constructed, one for each of the target serotypes considered (6A, 6B, 14, 19F, 23F). Thus, the carriage status of each study participant was recorded at each monthly visit as 0 (noncarrier), 1 (carrier of the target serotype T), 2 (carrier of any other serotype O), or 9 when either the swab was not taken or the laboratory result was not reported. The household (or family) state at each visit was then derived by combining the carriage status of all family members, resulting in a sequence of 0’s, 1’s, and 2’s if their information was complete (i.e., no missing values). For example, the nine possible states of a family of two individuals would be 00, 01, 02, 10, 11, 12, 20, 21, and 22.

The number of transitions between each pair of household states over a 28-day period was tabulated for each family. One household of seven was excluded from the analysis because their sequences were noninformative. Only complete family transitions, where the infection state of all household members was known on two consecutive observations, were used in the following analysis.

**Model**

The model by Melegaro et al. (20) was extended to incorporate serotype-specific information. In this first-order Markov model (or transition model) (23), the carriage status of each family at time *t* depends on the carriage status of all family members at time *t − 1*.

**Individual transitions.** The model assumes that individuals are not colonized with more than one pneumococcal serotype; if a carrier of one serotype is challenged by another serotype, then one type will rapidly outcompete the other. This process allows the model structure to remain relatively simple (three possible states to deal with) and close to our original model. The additional complication of a fourth state of co-colonized individuals (i.e., *C*12) within the modeling framework was not justified because there were only six occasions in the data set (0.16 percent) in which simultaneous carriage of more than one serotype was detected (21).

A susceptible-infected-susceptible structure (24) was assumed (figure 1); thus, no increased immunity after
a carriage episode was considered in the current model. The probability of a susceptible individual acquiring pneumococcus is defined as the sum of two components: the within-family acquisition rate and the acquisition rate from the community. The community acquisition rates are age and serotype specific (but constant over time and across different households), whereas the within-family acquisition rate is time dependent, being a function of the number of other household members carrying that particular serotype at that time. When individuals recover from a carriage episode (at an age- and serotype-dependent rate \( \mu \)), they again become susceptible to that specific type as well as to other types. Moreover, a new type can be acquired although the host is already colonized by a different pneumococcal serotype (\( C_1 \rightarrow C_2 \) and \( C_2 \rightarrow C_1 \)). The force of infection, however, for both of these flows is adjusted by two competition parameters (\( d_i \), where \( s = 1,2 \)) that provide the relative risks of acquiring serotype \( s \) if already colonized with the other type, compared with a noncarrier (25). In other words, \( d_1 \) and \( d_2 \) can also be considered two serotype-specific attributes, which we describe, respectively, as the challenging strength (i.e., the ability of one serotype to colonize when another serotype is present) and the susceptibility level (i.e., the ability of one serotype to remain colonizing when another serotype is introduced as a challenge).

This general competition model has two noteworthy special cases:

1. Complete exclusion (\( d_1 = d_2 = 0 \)): In this scenario, a resident serotype colonizing the host is always able to exclude challengers. The carriage episode may resolve at an age- and serotype-specific rate, but this rate is not affected by challenges from other serotypes.
2. Full susceptibility (\( d_1 = d_2 = 1 \)): In this scenario, any resident serotype is fully susceptible to challenges from other serotypes. A challenging serotype will terminate the carriage episode of the resident serotype and take over as the new resident serotype.

The simple age stratification (0–4 years, ≥5 years) used in our previous work was maintained (20). The transition probabilities between states in a short time interval \( \delta t \) are thus defined for an individual in the age class \( i = a,c \) (where \( c = \) child (<5 years) and \( a = \) adult (≥5 years)).

The following set of equations describes the model probabilities of an individual in age group \( i \) making transitions between the states in a short time interval \( \delta t \):

\[
P_i(C_s \rightarrow S)_{\delta t} = \mu_{i,s} \times \delta t \tag{1}
\]

\[
P_i(S \rightarrow C_s)_{\delta t} = \lambda_{i,s} \times \delta t \tag{2}
\]

\[
P_i(C_s \rightarrow C_l)_{\delta t} = d_l \times \lambda_{i,1} \times \delta t \tag{3}
\]

\[
P_i(C_1 \rightarrow C_2)_{\delta t} = d_2 \times \lambda_{i,2} \times \delta t, \tag{4}
\]

where \( \mu_{i,s} \) and \( \lambda_{i,s} \) are, respectively, the spontaneous clearance rates and the force of infection of type \( s \) for age class \( i \).

The force of infection of type \( s \) for an individual in age group \( i \) depends on the infection status of other members of the household:

\[
\lambda_{i,s} = \left( k_{i,s} + \beta_{is,s} I_{s,i}(t) + \beta_{ic,s} I_{c,i}(t) \right) / (z - 1)^w, \tag{5}
\]

where \( k_{i,s} \) is the community acquisition rates of type \( s \) for age class \( i \) and \( z \) is the family size. \( I_{s,i}(t) \) and \( I_{c,i}(t) \) are the number of, respectively, infected adults and infected children (type \( s \) in the family. \( \beta_{is,s} \) is the transmission rate of type \( s \) from an infected to an uninfected individual and reflects both the infectiousness of an individual in age class \( j \) and the susceptibility of an individual in age class \( i \) (where \( i = c,a \)). As in our previous work, we included a density correction factor in the denominator to adjust for possible changes in the contact pattern within families of different sizes. In the current model, \( w \) was set to 1, given that the estimate derived in the previous work was very close to this level (\( w = 1.184, 95 \) percent confidence interval: 0.2, 2.2). The model parameters are described in table 1.

**Family transitions.** The probabilities of transition between different family states \( r \) and \( s \) (\( p_{rs} \)) in a short time interval \( \delta t \) were derived by using equations 1–5, assuming that, in a short interval, only one member of the family will change status (through either infection or recovery). The probability of transitions in which more than one individual changes status was therefore set to zero. The probability that the family does not change state is given by \( q_{rr} = 1 - \sum_{s \neq r} p_{rs} \). Here, as in our previous work, we set \( \delta t = 1 \) day. A transition matrix was thus calculated for each household \( i \) (\( T_{i,1} \)). An example for a family of two (one adult and one child) containing all these probabilities is provided in Web appendix 1. (This information is described in the first of three supplementary appendices; each is referred to as “Web appendix” in the text and is posted on the Journal’s website (http://aje.oupjournals.org/)).

For each family, the matrix of transition probabilities for a 28-day interval was then derived as \( T_{28,i} = (T_{1,i})^{28} \). Thus, all possible pathways between the two states were implicitly included.
Parameterization and fitting strategy

Maximum likelihood techniques were adopted to estimate the model parameters (Web appendix 2) (26–28). The profile likelihood method was used to derive confidence intervals (29). Four parameterizations were considered (table 2), with increasing flexibility, which were compared by using Akaike Information Criterion values (Akaike Information Criterion = model deviance + 2 × number of parameters estimated) (30). The lower the value of the Akaike Information Criterion, the better the fit. The model was programmed in Matlab 6.5 (The MathWorks Inc., Natick, Massachusetts).

Vaccine type and nonvaccine type analysis

An additional analysis was conducted after stratifying the serotypes into vaccine serotypes and nonvaccine serotypes. Carriage of any serotype contained in the seven-valent vaccine was coded 1, and carriage of any other type was coded 2. This parallel analysis checked whether potential differences in transmission parameters were detectable if serotypes were grouped together in this way.

RESULTS

The transmission characteristics of the five most prevalent pneumococcal serotypes were explored and compared with the less common ones. All the serotype-specific models (M2–M4) gave an improved model fit over the non-serotype-specific model (M1), suggesting the existence of serotype-specific behaviors (refer to Web appendix 3 for details of models’ estimates and fits). The best fitting model varied between serotype: model 2 (serotype-specific spontaneous clearance rates) for serotype 19F; model 3 (serotype-specific clearance rates and proportionality factor for serotype-specific transmission) for serotypes 6A and 6B, and model 4 (serotype-specific clearance and transmission rates) for serotype 14. For serotype 23F, the improvement was nonsignificant, indicating that this serotype has no distinctive pattern that differentiates it from the average behavior of all the other types.

In agreement with our previous findings (20), we found that for all serotypes, acquisition rates from the community were significantly higher in preschool children than in older individuals (figure 2). Significant differences in the spontaneous clearance rate were also found between both age categories and serotypes (figure 3). Estimates for the nontarget

* Thecommunity acquisition rate for the target serotype is necessarily assumed different from the one related to all the other serotypes together. M, model.
serotypes were consistent across the different analyses (regardless of the target serotype), with an average duration of carriage of 18.5 days (range: 16–20 days) among individuals ≥5 years of age and 53.4 days (range: 44–59 days) at 0–4 years of age. Serotype-specific estimates similarly showed a lower spontaneous clearance rate for children than for older individuals; this result was statistically significant for serotypes 6A, 6B, and 23F. The estimated duration of carriage (1/spontaneous clearance rate) for serotype 6B in children was significantly longer than for other serotypes, reaching 116 days (95 percent confidence interval: 71, 220).

Transmissibility within the household was also investigated. We found significantly higher household transmission rates for serotype 6A (\( \alpha = 3.73 \), 95 percent confidence interval: 2.30, 5.65) and significantly lower rates for serotype 6B (\( \alpha = 0.49 \), 95 percent confidence interval: 0.22, 0.88) than for all other types. For serotype 14, the simple scaling of all household transmission rates by a single factor \( \alpha \) (model 3) was not the best fit; this was given by a more general model with significantly higher transmission rates from children to adults than estimated for other types.

The maximum likelihood estimates and likelihood-based 90 percent confidence limits for the challenging strength (\( d_1 \)) and the resistance capacity (\( 1 - d_2 \)) for each serotype are shown in figure 4. At this level of significance, no points in the figure are included in all five confidence limits (at the 95 percent level, there are a few), which suggests that differences between serotypes exist. It is striking that the complete exclusion model (\( d_1 = d_2 = 0 \)) is discarded for all five serotypes, suggesting that this mechanism is not appropriate for competition between colonizing pneumococcal serotypes. On the other hand, the opposite extreme scenario (\( d_1 = d_2 = 1 \)) is included in 14 and 6A regions but not the other three, for which a more general model is needed. We found large variation in the challenging strength, with the relative risk for a carrier of another serotype (compared with a noncarrier) acquiring \( T \) ranging from 12 percent for 19F to 100 percent for 6A and 6B. Moreover, the model showed that 6B is the only serotype characterized by a high resistance capacity (\( 1 - d_2 = 65 \) percent, 95 percent confidence interval: 31 percent, 88 percent); that is, carriage of 6B significantly reduces individual susceptibility to other types.

Table 3 summarizes the characteristics of the target serotypes. Similarities between serotypes 19F and 23F were found, that is, both are average type, with carriage episodes that last approximately 50 days in children, acquisition rates from within and outside the household similar to the rates for all the other serotypes, and low invading and resistance levels. Serotypes 6A and 14 share high within-household transmissibility as well as high-invading capacity, although both these types have no resistance capacity to challenging types. A distinct behavior appears to characterize serotype 6B, which, from the current analysis, showed an exceptional competing capacity as well as an extremely long period of within-host survival.

The vaccine serotypes/nonvaccine serotypes analysis found small differences in the characteristics of the groups. The former were carried on average for 64 (95 percent confidence interval: 50, 83) and 20 (95 percent confidence interval: 16, 27) days among, respectively, those aged 0–4 and ≥5 years, and nonvaccine serotypes were carried for a shorter period: 40 (95 percent confidence interval: 30, 54) and 13 (95 percent confidence interval: 10, 17) days, respectively.

**DISCUSSION**

In this paper, we further developed a modeling framework for analyzing longitudinal family carriage studies by extending our original model (20) to incorporate serotype-specific
information at the individual level. This extension is essential considering previous evidence of the diversity in the dynamics of different serotypes of *S. pneumoniae* (31–33). To continue to ignore this heterogeneity and the potential for competition between strains would have been inappropriate, especially considering that these characteristics may provide the mechanism for the serotype replacement observed following implementation of a universal pneumococcal conjugate vaccination policy (25, 34, 35). Similarities and differences among the five most prevalent pneumococcal serotypes were assessed through their transmissibility, acquisition, and clearance rates as well as their competing strengths.

We found that serotypes 6A, 6B, and 23F all share a longer mean duration of carriage in children 0–4 years of age than in those aged ≥5 years, which is in agreement with our previous work, where we estimated a mean duration of carriage of 51 and 19 days, respectively, in the two age strata. In particular, serotype 6B presents a very distinctive pattern, with a mean duration of almost 4 months in children and only 11 days in those ≥5 years of age. Slightly higher estimates of the duration of carriage were found in a recent study by Sleeman et al. (36). However, these authors studied only very young children (<2 years of age), and the average length of a pneumococcal carriage episode is longer at younger ages (33, 37, 38).

The serotype-specific properties we found with this analysis, related to both transmission and competition between different types, represent an important aspect in the analysis of host-pathogen interactions, and they highlight the existence of individual serotype-specific characteristics that would have been overlooked if we had exclusively considered the dynamics of vaccine serotypes versus nonvaccine serotypes. For example, we found only small and marginally significant differences between the duration of carriage for vaccine serotypes and nonvaccine serotypes groups. Similarly, recent work by Cauchemez et al. (39) found no significant difference in the mean duration of carriage of vaccine serotypes and nonvaccine serotypes groups.

Interestingly, the sum of the five serotype-specific community acquisition rates derived here (from the baseline model) is considerably higher than the overall community acquisition rate estimated in our non-serotype-specific analysis (0.019 vs. 0.012) (20). This finding highlights the main limitation of our previous work, which did not use the serotype information and left the model free to attribute a single prolonged carriage episode on occasions when carriage switched from one serotype to another.

Many characteristics of some serotypes were similar. Most notably, serotypes 19F and 23F were similar in terms of duration of carriage, within-household transmission, community acquisition rates, and low competing attributes. Serotypes 6A and 6B had similarly strong competition parameters, but 6B had a long mean duration of carriage, whereas 6A had a high transmissibility within the household. Serotype 14 also had high transmissibility within the household but weaker competition parameters than 6A. Characterizing nonvaccine serotypes in this way may help to identify which serotypes could replace vaccine serotypes after the introduction of vaccination. High competition strength will be less important when there are fewer competing serotypes; serotypes with low competition strength but long carriage duration and/or high transmissibility may be most likely to emerge as replacements. This would be a concern if they were also highly pathogenic.

Modeling work on interference between viruses was addressed by the pioneering works of Dietz (40) and Elveback et al. (41, 42), who looked at competition between two virus populations in a community and derived mathematical conditions for coexistence. Their work has been greatly extended in recent years (43–48), when several models for multistrain infections have addressed issues such as intra-specific competition, replacement, cross immunity between strains, and superinfection. We explored different mechanisms for serotype competition by investigating serotype-specific properties such as challenging and resistance strength. To our knowledge, this is the first time that the level of competition/interaction between different variants of the same organism in the human population has been estimated. Lipsitch et al. (35) investigated competition between two pneumococcal serotypes in a mouse model and found that carriage of the resident strain (6B) significantly inhibited acquisition of the challenge strain (23F) at the lowest challenge dose. This finding is in agreement with ours, which showed a high resistance capacity for serotype 6B as well as a low challenging capacity for serotype 23F.

Serotype replacement with nonvaccine serotypes has been observed for both pneumococcal carriage and invasive

---

**TABLE 3. Summary of the characteristics of the five most prevalent serotypes estimated by using a longitudinal family data set from the United Kingdom, October 2001–July 2002**

<table>
<thead>
<tr>
<th>Target serotype</th>
<th>Carriage duration ((1/\mu_c))</th>
<th>Transmissibility in the household</th>
<th>Acquisition from the community ((k_c))</th>
<th>Challenging strength ((d_l))</th>
<th>Resistance ((1 - d_l))</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>6B</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>6A</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>19F</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>23F</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The authors thank the European Union (QLG4-CT-2000-00640) for financial assistance.

The authors also thank Drs. Elisabeth Miller and Robert George for providing the pneumococcal carriage data and for very useful discussions on the manuscript. They also thank Dr. Mahein Hussain for organizing the study. They are grateful to all the nurses for keeping the study up and running and assisting the families involved throughout the study.

Conflict of interest: none declared.

REFERENCES