

Published in final edited form as:

Epidemiology. 2012 July ; 23(4): 510–519. doi:10.1097/EDE.0b013e31824f2f32.

Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model

Marc Lipsitch^{1,2}, Osman Abdullahi^{2,3}, Alexander D'Amour⁴, Wen Xie², Daniel M. Weinberger^{2,5}, Eric Tchetgen Tchetgen^{1,6}, and J. Anthony G. Scott^{3,7}

¹Center for Communicable Disease Dynamics and Department of Epidemiology, Harvard School of Public Health, Boston, MA USA

²Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA USA

³Kenya Medical Research Institute – Wellcome Trust Research Programme, Kilifi, Kenya

⁴Department of Statistics, Harvard University, Cambridge, MA

⁵Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

⁶Department of Biostatistics, Harvard School of Public Health, Boston, MA USA

⁷Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom

Abstract

BACKGROUND—There are more than 90 serotypes of *Streptococcus pneumoniae*, with varying biologic and epidemiologic properties. Animal studies suggest that carriage induces an acquired immune response that reduces duration of colonization in a non-serotype-specific fashion.

METHODS—We studied pneumococcal nasopharyngeal carriage longitudinally in Kenyan children aged 3–59 months, following up positive swabs at days 2, 4, 8, 16, and 32, and then monthly thereafter until two swabs were negative for the original serotype. As previously reported, 1868/2840 (66%) of children swabbed at baseline were positive. We estimated acquisition, clearance and competition parameters for 27 serotypes using a Markov transition model.

RESULTS—Point estimates of type-specific acquisition rates ranged from 0.00025/day (type 1) to 0.0031/day (type 19F). Point estimates of time to clearance (inverse of type-specific immune clearance rate) ranged from 28 days (type 20) to 124 days (type 6A). For the serotype most resistant to competition (type 19F), acquisition of other serotypes was 52% less likely (95% confidence interval= 37%–63%) than in an uncolonized host. Fitness components (carriage duration, acquisition rate, lack of susceptibility to competition) were positively correlated with each other and with baseline prevalence, and were associated with biologic properties previously shown to associate with serotype. Duration of carriage declined with age for most serotypes.

CONCLUSIONS—Common *S. pneumoniae* serotypes appear superior in many dimensions of fitness. Differences in rate of immune clearance are attenuated as children age and become capable of more rapid clearance of the longest-lived serotypes. These findings provide information for comparison after introduction of pneumococcal conjugate vaccine.

Streptococcus pneumoniae is a bacterium that commonly colonizes the nasopharynx of children, causing a range of diseases when it invades normally sterile sites. Pneumococcal disease causes an estimated 11% of deaths in children aged under 5 years worldwide¹ and considerable morbidity and mortality in adults.^{2–5} The nasopharynx is the site from which

pneumococci are ordinarily transmitted,⁶ and nasopharyngeal carriage is thought to precede disease.⁶ Over 90 pneumococcal serotypes – defined by the polysaccharide capsule – are known.⁷ Capsular variation has important consequences for many aspects of the host-pathogen interaction.⁸⁻¹⁵

Several immune mechanisms protect against nasopharyngeal carriage, thereby reducing disease risk and disrupting transmission. Licensed pneumococcal conjugate vaccines confer serotype-specific, antibody-mediated immunity, which reduces nasopharyngeal carriage,¹⁶ invasive disease¹⁷ and mortality.¹⁸ For some serotypes, there is evidence that naturally acquired anticapsular antibody can also protect against carriage.¹⁹⁻²¹ There is also acquired immunity to colonization that is not serotype-specific.²² A likely mechanism of this immunity²³ is an antigen-specific CD4⁺ T cell response, which acts via interleukin-17A secretion and neutrophil recruitment to hasten clearance of colonization in older (more immune) hosts.^{24,25} Some next-generation pneumococcal vaccines are designed to elicit this form of immunity.²⁶

Longitudinal studies are required to assess how the development of immunity to pneumococci in children determines the age- and serotype-specific pattern of carriage and disease.²⁷⁻³² Knowledge of this process is necessary to improve our understanding of pneumococcal diversity in unvaccinated populations³³ and the impact of vaccines on this diversity and resulting patterns of disease.^{13,34,35} Determinants of serotype frequency include rates of acquisition of carriage and clearance of carriage, as well as the strength of competitive interactions that determine whether a subsequent serotype can colonize a person already colonized with one serotype.^{27,30,36,37}

A large longitudinal study of pneumococcal carriage undertaken in children 3-59 months old in Kilifi, Kenya^{38,39} offers an opportunity to estimate these parameters. Previously, we estimated acquisition and loss rates from this study, using a separate model for each serotype and assuming constant hazards of acquisition and loss.³⁹ Here we extend this work using a single, more detailed model that includes states for colonization with each of the 27 most common serotypes (collectively accounting for 95% of all baseline colonization, plus a category for “other” serotypes) and simultaneously estimates acquisition and clearance rates of each serotype, as well as a parameter measuring each serotype’s competitive ability. We also examine age-related changes in serotype-specific parameters. Questions motivating this analysis were: (1) How do serotypes differ in acquisition rates, immune clearance rates, and susceptibility to competitive encroachment by other types? (2) If such differences are found, are they consistent with *in vitro* measures of capsular structure and function that have previously been found to predict serotype carriage prevalence?¹⁵ (3) For each serotype, does time to clearance of pneumococcal carriage decrease with age, consistent with animal experimental studies of immune maturation⁴⁰ and acquired immunity following exposure to carriage?²⁵

Methods

Study population

The study was conducted among residents of the Kilifi Health and Demographic Surveillance Study area, Kilifi, Kenya. This study is a longitudinal recording of residents in a well-defined geographic area (population about 250,000) around Kilifi District Hospital on the Kenyan coast, updated through continuous census for vital events every four months. The study was approved by the Kenya Medical Research Institute/National Ethical Review Committee and The Oxford Tropical Research Ethics Committee, and written informed consent was obtained for all participants.

Study design

The study has been described elsewhere.^{38,39} Briefly, we randomly selected 3570 children aged 3-59 months from the surveillance study register and approached their parents/guardians for consent between October 2006 and November 2008. Exclusion criteria included migration or planned migration from the study area, lack of parent/guardian consent, or illness that prevented the taking of a nasopharyngeal swab. Children whose day-1 (“baseline”) swab cultures were positive for pneumococcus were resampled on approximately days 2, 4, 8, 16, and 32 and monthly thereafter, until obtaining two consecutive swabs, each of which had either no pneumococcal growth or a pneumococcal serotype different from the baseline serotype. Once 50 complete durations had been documented for a serotype, additional children found with that serotype at baseline were not followed up, but baseline prevalence estimates included individuals beyond the 50-person limit. As previously reported,^{38,39} the study contained baseline swabs on 2840 children, of whom 1868 (66%) carried pneumococci at baseline.

Data Analysis

We defined a Markov transition model with 29 states, corresponding to uncolonized with *S. pneumoniae* (state 0), colonized with each of the 27 serotypes most commonly recovered in baseline samples (states 1-27), or colonized with any other serotype (state 28). For each “serotype” (including other) i in the model, the transitions within this Markov model are: acquisition (transition from uncolonized state 0 to state i , occurring with hazard rate a_i), clearance (transition from state i to state 0, occurring with hazard rate b_i), and change of serotype (transition from state i , which we call “resident,” to state j , which we call “challenge,” occurring with hazard rate c_{ij}). All hazards were assumed to be time-homogeneous and in the basic model identical for all hosts. Figure 1 shows a simplified version of this model (with two individual serotypes, labeled 1 and 2, and an “other serotype” category). Table 1 summarizes these model parameters.

In this study, the processes of clearance and change of serotype were observed (though interval-censored). We assumed that, for a child who completed followup (2 consecutive visits with other than the baseline serotype), (1) the first of these two visits indicated the state to which they switched after the baseline state, (2) the change occurred sometime between the last visit positive for baseline type and the first of the two consecutive visits not positive for it, and (3) no other changes occurred unobserved during the period of followup. We also assumed that, at most, one event occurred in an interval between observations. Thus the child’s likelihood contribution was similar to that of an interval-censored observation in a competing risks setting (where the competing risks are clearance or switching to each of the other types), while children who are censored before two consecutive visits negative for the baseline serotype are assumed to have been censored at the time of the last visit.

Acquisition from the uncolonized state was never observed, because subjects uncolonized at baseline are censored after their first visit. Instead, we estimated acquisition rates by assuming that the baseline sample represents a sample from the stationary distribution of the Markov process, which is a deterministic (but algebraically complicated) function of the hazards.

For prevalence estimation in the baseline sample, we included all 2840 children, indexed by x . For estimating rates of clearance and switches to other serotypes, we included only those children who were colonized at baseline (denoted $o_x = 0$) and from whom at least 3 swabs were taken in total (such that two swabs negative for the baseline type might be observed, denoted $y_x = 1$).

Thus we can compute a likelihood for the data on each individual x with baseline state i_x , ending state j_x , time of sampling t_x for the last swab positive for type i before the two negatives, time of sampling u_x for the first of two consecutive swabs negative for type i_x or (if this does not occur), the last swab positive for type i_x ; indicator $q_x = 1$ if individual x was still carrying i_x at time u_x and $q_x = 0$ otherwise, and indicator $m_x = 1$ if individual x had cleared carriage at time u_x , and $m_x = 0$ otherwise. Suppressing the subscript x for readability, each child makes an independent likelihood contribution \bar{p}_0 if uncolonized at baseline and \bar{p}_i if carrying type i at baseline, where \bar{p}_i is the stationary probability (a function of the a_i, b_i, c_{ij}) of being in state i (including uncolonized, $i = 0$).

In addition, for children positive at baseline who receive at least 3 swabs, those in whom no

transition has occurred by time u contribute likelihood $e^{-h_i u}$, where $h_i = b_i + \sum_{j \neq i} c_{ij}$ is the total hazard of moving away from state i . Those who clear colonization between t and u

contribute likelihood $\frac{b_i}{h_i} (e^{-h_i t} - e^{-h_i u})$, while those who acquire type j between t and u

contribute likelihood $\frac{c_{ij}}{h_i} (e^{-h_i t} - e^{-h_i u})$. Overall the likelihood contribution for individual x is

$$L_x(\theta; i, j, q, m, u, t, o) = \frac{o^{1-o}}{p_0 p_i} \left\{ \left[\frac{b_i^m c_{ij}^{1-m}}{h_i} (e^{-h_i t} - e^{-h_i u}) \right]^{1-q} e^{-h_i q u} \right\}^{y(1-o)}$$

(See Table 2.) The likelihood of the full data is the product of this likelihood over all subjects.

For n serotypes, such a model would have $n^2 + n$ independent rates, more than are feasible to fit even with this large data set. To reduce dimensionality, we considered four models, each of which independently estimates the $2n$ acquisition and clearance rates, but constrains the “change of serotype” rates c_{ij} in various ways.

- A. “Fixed susceptibility to competition” in which the acquisition rate of type j among children colonized with type i is a fixed constant r (susceptibility to competition) times the acquisition rate among uncolonized persons: $c_{ij} = r a_j$
- B. “Type-specific susceptibility to competition” in which the acquisition rate of type j among children colonized with type i is a resident-type-specific susceptibility to competition r_i times the acquisition rate among uncolonized persons: $c_{ij} = r_i a_j$
- C. “Type-specific challenge strength” in which the acquisition rate of type j among children colonized with type i is specific to type j , but independent of the serotype i that is resident and not necessarily proportional to the acquisition rate of j among uncolonized children: $c_{ij} = s_j$
- D. “Full model” (combining B and C) in which the acquisition rate of type j among children colonized with type i is a resident-type-specific constant r_i times a challenge-type-specific constant s_j : $c_{ij} = r_i s_j$

These models are summarized in Table 3. We fit each model by maximum likelihood using the optim routine, BFGS and L-BFGS-B methods, in R 2.11.1,⁴¹ ranked them by Akaike information criterion (AIC), and assessed each for face validity and stability of estimates.

Correlations of serotype characteristics measured in this study and in prior studies

Pearson correlations among various properties of serotypes from this study were estimated. In these correlations, prevalence was logit-transformed; acquisition and loss rates, and susceptibility to competition were log-transformed.⁴²

We assessed the relationship of specific laboratory-measured characteristics of serotypes (number of carbons per polysaccharide repeat unit, a measure of capsule complexity and predictor of capsule thickness; and survival of the serotype in a standardized genetic background when exposed to human neutrophils in a surface killing assay¹⁵) to the estimated rates from model B by Spearman rank correlation coefficients.

Confidence intervals for Pearson correlations were calculated using the delta method.⁴³ We transformed the model parameters into the arctanh of the relevant correlations, and used the gradient of this transformation to perturb the asymptotic distribution of the parameters' maximum likelihood estimates. We used the arctanh transformation, following Fisher's variance stabilizing z-transformation,⁴⁴ to calculate the confidence intervals on a scale where the normal approximation would be appropriate. Confidence intervals in the arctanh scale were then mapped back to the standard correlation scale.

P-values for Spearman correlations were calculated using the asymptotic t-test statistic for the permutation test.⁴⁵

Effect of age on serotype-specific measures

We repeated parameter estimates within approximate terciles of age, with age cutoffs at 22 and 41 months. Because of data sparsity, we considered the 23 most common serotypes at baseline (rather than the 27 serotypes we used in the all-ages analysis) and combined the remaining ones in an "other" category. Because data for some serotypes in some age groups were sparse, we used the asymptotically normal sampling distribution of the all-ages estimates to construct weak priors for each of the age-specific analyses.

Detailed analyses of longitudinal samples

We compared the frequency of each serotype in baseline samples to that in follow-up samples that were different from baseline. We also assessed whether particular serotypes appeared together at adjacent time-points in a given child more often than expected from their population frequencies (e.g., whether particular serotypes tended to colonize particular hosts or to be better at "sharing" a host). We did so first by identifying every "transition pair" of adjacent samples from a single host in which pneumococci of different serotypes were present in the two samples, and noting the starting and ending serotype in each pair, creating a histogram of the number of pairs that (for example) started with type 19F and ended with 23F. We then created 400 permutations of these pairs (permuting the ending serotype) and obtained an expected value and distribution of the number of pairs with each start and end serotype. After correcting for the appearance of pairs with the same start and end serotype in the permutations, we calculated the observed-minus-expected number of pairs for each pair of serotypes, and created a smoothed curve of the 95th percentile of the observed-expected statistic as a function of the number of expected pairs. Serotype pairs falling outside this 95th percentile were examined.

RESULTS

Details of the frequencies and types of observations used in the model are shown in eTable 1 (<http://links.lww.com/EDE/A581>).

Model selection

Maximum likelihood estimates converged for all four models (Table 1). The best-fitting models by AIC were models B (with type-specific susceptibility to competition) and D (the full model). However, model B provided stable estimates and finite confidence intervals for all parameters for all 28 serotypes, while the estimates in model D were extremely unstable, including confidence intervals spanning zero to infinity, even for quantities associated with some common serotypes, such as 19F and 23B. For this reason, we proceeded to analyze only model B.

Time to clearance

Figure 2A shows the estimated time to clearance by serotype (reciprocal of clearance rate), showing serotypes in declining order of baseline prevalence. Point estimates of time to clearance ranged from 28 days for type 20 to 123 days for type 6A. .

Acquisition rate among uncolonized persons

Figure 2B shows the estimated rates of acquisition by serotype, preserving the same order of serotypes. Point estimates of acquisition rates ranged from 0.00025/day for type 1 to 0.0031/day for type 19F.

Susceptibility to competition

Figure 2C shows serotype-specific susceptibility to competition, defined as the rate at which individuals carrying that serotype switch to carry another serotype, relative to the rate at which that other serotype colonizes an uncolonized person $r_i = c_{ij} / a_j$. Point estimates ranged from 0.48 (serotype 19F) to 2.1 (serotype 33B). While confidence intervals for r_i for many serotypes are strictly below 1.0 (giving evidence for competition), the serotypes with point estimates above 1.0 are mainly the rarest ones, most of which have confidence intervals that dip below 1.0. Thus, these results are consistent with the presence of competition by all strains, though its strength is uncertain for the rarer types.

Correlations among fitness components and with laboratory measurements

Table 4 shows the correlation coefficients among prevalence, acquisition rate, clearance rate and susceptibility to competition. The first four properties are correlated with each other in such a way that the most-“fit” serotypes on each property tend to do best on the others too: serotypes with higher prevalence tend to have lower hazards of clearance, are less susceptible to competition, and have higher rates of acquisition

We also tested the hypothesis that these epidemiologic properties of serotypes would correlate with two additional properties of serotypes that we have previously studied: the biochemical complexity of the capsular polysaccharide repeat unit (measured as number of carbons per repeat unit) and in vitro resistance to nonopsonic surface phagocytosis by human neutrophils.¹⁵ For the subset of serotypes for which these measures are available, such correlations are suggested with each of the studied epidemiologic properties. Greater resistance to surface phagocytosis is associated with higher acquisition rate, lower clearance rate and better competitive ability; to a lesser degree, simpler capsular polysaccharide repeat units are also associated with these properties.

Changes in rates and susceptibility to competition with age

Time to clearance became faster with age for virtually every serotype (Figure 3A). The serotype-specific time to immune clearance fell from a median (interquartile range, IQR) of 105 days (79-137) in those under 22 months to 39 days (30-43) in those 41 months or older. The estimated time to clearance of a serotype among children <22 months was nearly

uncorrelated with that in the two older age groups, suggesting that the factors predicting duration of colonization in the younger age groups become less important in older children. The serotypes with the longest time to clearance in young children had the greatest proportional decline in time to clearance (Figure 4) (log values, Pearson $r=-0.86$, [95% CI=-0.71 to -0.94]).

Acquisition rates showed no obvious trend with age (Figure 3B). For several serotypes, incidence in uncolonized children increased with age, suggesting that increasing exposure of older children to colonization may overcome any effect of serotype-specific immunity on age-specific incidence. Similarly, susceptibility to competition showed no consistent age pattern (Figure 3C).

Baseline vs. follow-up samples

Figure 4 shows the ratio of the prevalence of a serotype among baseline samples to its appearance (as a new serotype) among follow-up samples, plotted against its baseline frequency. This distribution is approximately symmetric around 1 for the rare to intermediate serotypes, but the 14 most common serotypes at baseline tended to be more common among baseline samples. This is unsurprising in that the types most common at baseline are those with high incidence, long duration, and low susceptibility to competition (Table 2), while presence at follow-up reflects only high incidence, and hence does not as strongly favor the high-prevalence types.

Co-occurrence of serotypes

Of the 529 pairs of serotypes considered, 39 (7.4%) occurred more often than the 95% cutoff estimated from permutations, suggesting they might be unusually common (eTable 2). Of these, 14 (including 10 of the top 15) were pairs of the same serogroup, which might be expected to occur due to serotyping errors: an international validation study found that approximately 5% of serotypes are erroneous, mainly due to assignment of the wrong type within a serogroup.⁴⁶ These co-occurrences may also reflect the fact that small genetic changes can lead to expression of another serotype within a serogroup (e.g. 15B-C, which account for the strongest two outliers in this analysis, differ by a point mutation, and often “switch” in carriage studies^{47,48}). The remaining 25 (4.7%) are consistent with the 5% that should occur by chance. These data show no evidence that certain pairs of serotypes tend to occur together in the same host.

DISCUSSION

The Kilifi longitudinal carriage study represents the largest and most intensively sampled data set to our knowledge for estimating basic parameters of pneumococcal acquisition, clearance and competition. We have here applied a relatively high-dimensional Markov model to estimate these parameters, taking advantage of the scale of the dataset while maintaining some parametric assumptions that allow for simultaneous estimation of all of the parameters of interest.

A previous analysis of this data set³⁹ estimated, separately for each serotype, rates of acquisition and loss (clearance or replacement by another serotype), using a statistical model with only two states – (1) colonized with a focal serotype or (2) uncolonized/colonized with another serotype. That approach was much simpler and gave acquisition- and clearance- rate estimates strongly correlated with, but different from, those estimated here.

The more complex approach used here overcomes three limitations of the previous estimates. First, we separately estimate rates of immune clearance and rates of replacement by other serotypes. Second, the previous method violated the Markov assumption by

combining very heterogeneous serotypes into the “other” category in the model for each serotype, which is less of a problem here with more states. Third, use of a single model here allows estimation of covariances in the different parameter estimates, aiding statistical analysis.

Even this model relies on simplifying assumptions that are violated in reality. For example, age strongly affects time to clearance, whereas a simple Markov model assumes that hosts are homogeneous. While the age-specific analysis accounts for the major effects of age, it still leaves out further sources of inter-individual heterogeneity in rates, such as exposure to others who are colonized, antimicrobial use, immune status unrelated to age, and others. We assume a constant hazard of each transition (acquisition, replacement with a new type or clearance of the current type), which may be violated in reality. Antimicrobial use is not documented in this study. Because serotypes vary in their degree of antimicrobial resistance,⁴⁹ clearance by antimicrobial use will reduce the duration more in those serotypes that are more susceptible. Our estimates reflect the local patterns of antimicrobial use and serotype-specific resistance; the latter are relatively similar to those in other settings prior to conjugate vaccine introduction.⁴⁹⁻⁵¹

Importantly, we do not explicitly account for simultaneous carriage of multiple serotypes, which cannot be detected by serotyping a single bacterial colony, as was the standard practice in this study. This omission may bias estimates of the competition parameter, though the direction of bias is unclear: observed changes in serotype may be due to changing detection of the already-resident serotype (thus inflating rates of acquisition of new serotypes), but acquisition events may also be missed (thus reducing rates of acquisition of new types). Duration estimates will mainly be biased downward by failure to detect a serotype when another is present, although the requirement of two swabs negative for the baseline type should mitigate this bias. Prevalence estimates will be biased downward by failure to detect multiple colonization.

Despite these limitations, the results show a clear pattern. Highly prevalent serotypes have the longest time to clearance, the lowest susceptibility to being replaced by other types, and the highest incidence. From the perspective of bacterial population biology, these components of “fitness” tend to be positively correlated, suggesting that serotype has a major impact on many aspects of the bacteria’s ability to spread and persist, and that those with higher fitness along one dimension tend to be fitter on other dimensions as well. Interestingly, a laboratory measure previously reported to correlate positively with prevalence – greater resistance to nonopsonic surface killing by human neutrophils¹⁵ – also correlates with these other fitness measures. While prevalence should correlate with these other components (since it is determined by them), further work is needed to understand other causal links among these laboratory and epidemiologic measures. While some serotypes reduced detectable acquisition of other strains, there was little statistical evidence that carriage of any serotype facilitates acquisition of others. Biologically, it is expected that presence of a resident strain might inhibit acquisition of another strain,^{27,30,36,37} but there is no known mechanism by which the presence of a resident strain could promote acquisition of another strain.

We also find that, for many serotypes (especially the longest-lived), time to clearance decreases with age. This is consistent with earlier findings.^{24,30} In our study, by the time children reach 41 months of age, the mean time to clearance is on the order of 1-3 months for most serotypes (less than half that for children <22 months), and time to clearance in these older children is no longer so closely correlated with other components of fitness. Reduced duration of carriage with age for many serotypes is consistent with the expectation from animal studies of the CD4+ T-cell dependent responses acquired after exposure to

pneumococcal carriage, which reduce the duration of pneumococci via immune responses not targeted at the capsular serotype.^{23,25} Reduced duration of carriage with age is also consistent with a maturation of the immune response with age, independent of exposure, which in another animal study was shown to hasten pneumococcal clearance.⁴⁰

Several longitudinal studies have estimated acquisition and loss parameters for pneumococcal serotypes. Precise comparison across studies is difficult because other studies have estimated acquisition rates conditional on the presence of a colonized person in the family^{29,31,32} or school,²⁸ or because serotypes were grouped into heterogeneous groups.²⁸ Bangladeshi infants³² cleared common serotypes somewhat faster (approximately 30-40 days) than estimated in our study. Serotype-specific carriage durations (1-2 months) in Danish children (median age 1.9 years at enrollment) in day care were more similar to those observed here for children older than 41 months of age, perhaps because the higher prevalence in Danish day-care attendees had stimulated more rapid development of immunity.²⁷ Also, colonization was more protective (approximately 90%) against acquisition of a new serotype in that study than here.²⁷ These studies differed in sampling times, consideration of multiple carriage, and setting.

Precise estimates of serotype-specific epidemiologic parameters are important to understand coexistence of different serotypes^{33,52} and to understand the impact of vaccination on serotype dynamics.^{13,37,53} Future studies should estimate similar parameters for the serotypes that are rare in unvaccinated populations but become common following vaccination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Footnotes

All other authors declare no conflict of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the author.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Acknowledgments

We thank Sarah Cobey and James Nokes for helpful comments.

ML has received consulting fees or honoraria from Pfizer, Novartis, Outcome Sciences, and AIR Worldwide. JAGS has received grant funding from GlaxoSmithKline and has received study vaccines without cost from Pfizer/Wyeth.

FUNDING: This work was supported by research grant R01 AI04893 from the US National Institutes of Health to ML and by a clinical fellowship from the Wellcome Trust of Great Britain (No. 081835) to JAGS. This paper is published with the permission of the Director of the Kenya Medical Research Institute

REFERENCES

1. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*. 2009; 374(9693):893–902. [PubMed: 19748398]
2. Evers SM, Ament AJ, Colombo GL, Konradsen HB, Reinert RR, Sauerland D, Wittrup-Jensen K, Loiseau C, Fedson DS. Cost-effectiveness of pneumococcal vaccination for prevention of invasive pneumococcal disease in the elderly: an update for 10 Western European countries. *Eur J Clin Microbiol Infect Dis*. 2007; 26(8):531–40. [PubMed: 17570001]
3. Fedson DS, Scott JA. The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known. *Vaccine*. 1999; 17(Suppl 1):S11–8. [PubMed: 10471174]
4. Feikin DR, Jagero G, Aura B, Bigogo GM, Oundo J, Beall BW, Karani A, Morpeth S, Njenga MK, Breiman RF. High rate of pneumococcal bacteremia in a prospective cohort of older children and adults in an area of high HIV prevalence in rural western Kenya. *BMC Infect Dis*. 2010; 10:186. [PubMed: 20573224]
5. Scott JA, Hall AJ, Muyodi C, Lowe B, Ross M, Chohan B, Mandaliya K, Getambu E, Gleeson F, Drobniewski F, Marsh K. Aetiology, outcome, and risk factors for mortality among adults with acute pneumonia in Kenya. *Lancet*. 2000; 355(9211):1225–30. [PubMed: 10770305]
6. Bogaert D, De Groot R, Hermans PW. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis*. 2004; 4(3):144–54. [PubMed: 14998500]
7. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet*. 2006; 2(3):e31. [PubMed: 16532061]
8. Battig P, Hathaway LJ, Hofer S, Muhlemann K. Serotype-specific invasiveness and colonization prevalence in *Streptococcus pneumoniae* correlate with the lag phase during in vitro growth. *Microbes Infect*. 2006; 8(11):2612–7. [PubMed: 16938479]
9. Battig P, Muhlemann K. Capsule genes of *Streptococcus pneumoniae* influence growth in vitro. *FEMS Immunol Med Microbiol*. 2007; 50(3):324–9. [PubMed: 17578523]
10. Brueggemann AB, Peto TE, Crook DW, Butler JC, Kristinsson KG, Spratt BG. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *J Infect Dis*. 2004; 190(7):1203–11. [PubMed: 15346329]
11. Hanage WP, Auranen K, Syrjanen R, Herva E, Makela PH, Kilpi T, Spratt BG. Ability of pneumococcal serotypes and clones to cause acute otitis media: implications for the prevention of otitis media by conjugate vaccines. *Infect Immun*. 2004; 72(1):76–81. [PubMed: 14688083]
12. Hanage WP, Kajjalainen TH, Syrjanen RK, Auranen K, Leinonen M, Makela PH, Spratt BG. Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. *Infect Immun*. 2005; 73(1):431–5. [PubMed: 15618181]
13. Weinberger D, Harboe ZB, Flasche S, Scott JA, Lipsitch M. Prediction of serotypes causing invasive pneumococcal disease in unvaccinated and vaccinated populations. *Epidemiology*. 2011; 22:199–207. [PubMed: 21646962]
14. Weinberger DM, Harboe ZB, Sanders EA, Ndiritu M, Klugman KP, Ruckinger S, Dagan R, Adegbola R, Cutts F, Johnson HL, O'Brien KL, Anthony Scott J, Lipsitch M. Association of serotype with risk of death due to pneumococcal pneumonia: a meta-analysis. *Clin Infect Dis*. 2010; 51(6):692–9. [PubMed: 20715907]
15. Weinberger DM, Trzcinski K, Lu YJ, Bogaert D, Brandes A, Galagan J, Anderson PW, Malley R, Lipsitch M. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog*. 2009; 5(6):e1000476. [PubMed: 19521509]
16. Rinta-Kokko H, Dagan R, Givon-Lavi N, Auranen K. Estimation of vaccine efficacy against acquisition of pneumococcal carriage. *Vaccine*. 2009; 27(29):3831–7. [PubMed: 19490983]
17. Black S, Shinefield H, Fireman B, Lewis E, Ray P, Hansen JR, Elvin L, Ensor KM, Hackell J, Siber G, Malinoski F, Madore D, Chang I, Kohberger R, Watson W, Austrian R, Edwards K. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children.

- Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr Infect Dis J.* 2000; 19(3):187–95.
18. Cutts FT, Zaman SM, Enwere G, Jaffar S, Levine OS, Okoko JB, Oluwalana C, Vaughan A, Obaro SK, Leach A, McAdam KP, Biney E, Saaka M, Onwuchekwa U, Yallop F, Pierce NF, Greenwood BM, Adegbola RA. Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *Lancet.* 2005; 365(9465):1139–46. [PubMed: 15794968]
 19. Goldblatt D, Hussain M, Andrews N, Ashton L, Virta C, Melegaro A, Pebody R, George R, Soininen A, Edmunds J, Gay N, Kayhty H, Miller E. Antibody responses to nasopharyngeal carriage of *Streptococcus pneumoniae* in adults: a longitudinal household study. *J Infect Dis.* 2005; 192(3):387–93. [PubMed: 15995951]
 20. Granat SM, Mia Z, Ollgren J, Herva E, Das M, Piirainen L, Auranen K, Makela PH. Longitudinal study on pneumococcal carriage during the first year of life in Bangladesh. *Pediatr Infect Dis J.* 2007; 26(4):319–24. [PubMed: 17414395]
 21. Weinberger DM, Dagan R, Givon-Lavi N, Regev-Yochay G, Malley R, Lipsitch M. Epidemiologic evidence for serotype-specific acquired immunity to pneumococcal carriage. *J Infect Dis.* 2008; 197(11):1511–18. [PubMed: 18471062]
 22. Granat SM, Ollgren J, Herva E, Mia Z, Auranen K, Makela PH. Epidemiological evidence for serotype-independent acquired immunity to pneumococcal carriage. *J Infect Dis.* 2009; 200(1):99–106. [PubMed: 19469705]
 23. Zhang Q, Leong SC, McNamara PS, Mubarak A, Malley R, Finn A. Characterisation of regulatory T cells in nasal associated lymphoid tissue in children: relationships with pneumococcal colonization. *PLoS pathogens.* 2011; 7(8):e1002175. [PubMed: 21852948]
 24. Hogberg L, Geli P, Ringberg H, Melander E, Lipsitch M, Ekdahl K. Age- and serogroup-related differences in observed durations of nasopharyngeal carriage of penicillin-resistant pneumococci. *J Clin Microbiol.* 2007; 45(3):948–52. [PubMed: 17202280]
 25. Lu YJ, Gross J, Bogaert D, Finn A, Bagrade L, Zhang Q, Kolls JK, Srivastava A, Lundgren A, Forte S, Thompson CM, Harney KF, Anderson PW, Lipsitch M, Malley R. Interleukin-17A mediates acquired immunity to pneumococcal colonization. *PLoS Pathog.* 2008; 4(9):e1000159. [PubMed: 18802458]
 26. Moffitt KL, Malley R. Next generation pneumococcal vaccines. *Curr Opin Immunol.* 2011
 27. Auranen K, Mehtala J, Tanskanen A, M SK. Between-strain competition in acquisition and clearance of pneumococcal carriage--epidemiologic evidence from a longitudinal study of day-care children. *Am J Epidemiol.* 2010; 171(2):169–76. [PubMed: 19969530]
 28. Cauchemez S, Temime L, Valleron AJ, Varon E, Thomas G, Guillemot D, Boelle PY. S. pneumoniae transmission according to inclusion in conjugate vaccines: Bayesian analysis of a longitudinal follow-up in schools. *BMC Infect Dis.* 2006; 6:14. [PubMed: 16445857]
 29. Domenech de Celles M, Opatowski L, Salomon J, Varon E, Carbon C, Boelle PY, Guillemot D. Intrinsic epidemicity of *Streptococcus pneumoniae* depends on strain serotype and antibiotic susceptibility pattern. *Antimicrob Agents Chemother.* 2011; 55(11):5255–61. [PubMed: 21788454]
 30. Melegaro A, Choi Y, Pebody R, Gay N. Pneumococcal carriage in United Kingdom families: estimating serotype-specific transmission parameters from longitudinal data. *Am J Epidemiol.* 2007; 166(2):228–35. [PubMed: 17517684]
 31. Melegaro A, Gay NJ, Medley GF. Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiol Infect.* 2004; 132(3):433–41. [PubMed: 15188713]
 32. Erasto P, Hoti F, Granat SM, Mia Z, Makela PH, Auranen K. Modelling multi-type transmission of pneumococcal carriage in Bangladeshi families. *Epidemiol Infect.* 2010; 138(6):861–72. [PubMed: 20018129]
 33. Lipsitch M, O'Hagan JJ. Patterns of antigenic diversity and the mechanisms that maintain them. *J R Soc Interface.* 2007; 4(16):787–802. [PubMed: 17426010]
 34. Melegaro A, Choi YH, George R, Edmunds WJ, Miller E, Gay NJ. Dynamic models of pneumococcal carriage and the impact of the Heptavalent Pneumococcal Conjugate Vaccine on invasive pneumococcal disease. *BMC Infect Dis.* 2010; 10:90. [PubMed: 20377886]

35. Flasche S, Van Hoek AJ, Sheasby E, Waight P, Andrews N, Sheppard C, George R, Miller E. Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in England: a cross-sectional study. *PLoS Med*. 2011; 8(4):e1001017. [PubMed: 21483718]
36. Lipsitch M, Dykes JK, Johnson SE, Ades EW, King J, Briles DE, Carlone GM. Competition among streptococcus pneumoniae for intranasal colonization in a mouse model. *Vaccine*. 2000; 18(25):2895–901. [PubMed: 10812233]
37. Weinberger D, Malley R, Lipsitch M. Serotype replacement in disease following pneumococcal vaccination: A discussion of the evidence. *Lancet*. 2011 In press.
38. Abdullahi, O.; Karani, A.; Mugo, D.; Kungo, S.; Wanjiro, E.; Jomo, J.; Musyimi, R.; Nokes, J.; Tchetgen, E.; Lipsitch, M.; Scott, JA. The epidemiology of pneumococcal colonization in the nasopharynges of children in Kilifi District, Kenya. 2011. in preparation
39. Abdullahi O, Karani A, Mugo D, Kungo S, Wanjiro E, Jomo J, Musyimi R, Nokes J, Tchetgen E, Lipsitch M, Scott JA. The rates of clearance and acquisition of pneumococcal serotypes in the nasopharynges of children in Kilifi District, Kenya. *Journal of Infectious Diseases*. 2011 under review.
40. Bogaert D, Weinberger D, Thompson C, Lipsitch M, Malley R. Impaired innate and adaptive immunity to *Streptococcus pneumoniae* and its effect on colonization in an infant mouse model. *Infect Immun*. 2009; 77(4):1613–22. [PubMed: 19168741]
41. R Core Development Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. URL: <http://www.R-project.org>
42. Freeman J, Hutchison GB. Prevalence, incidence and duration. *American Journal of Epidemiology*. 1980; 112(5):707–23. [PubMed: 6969024]
43. Casella, G.; Berger, RL. *Statistical inference*. 2nd ed. Thomson Learning; Australia ; Pacific Grove, CA: 2002.
44. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*. 1915; 10(4):507–521.
45. Kendall, MG.; Stuart, A.; Ord, JK.; Arnold, SF.; O'Hagan, A. *Kendall's advanced theory of statistics*. 6th ed. Vol. 2. Edward Arnold ; Halsted Press; London New York: 1994.
46. Konradsen HB. Validation of serotyping of *Streptococcus pneumoniae* in Europe. *Vaccine*. 2005; 23(11):1368–73. [PubMed: 15661385]
47. Meats E, Brueggemann AB, Enright MC, Sleeman K, Griffiths DT, Crook DW, Spratt BG. Stability of serotypes during nasopharyngeal carriage of *Streptococcus pneumoniae*. *J Clin Microbiol*. 2003; 41(1):386–92. [PubMed: 12517877]
48. van Selm S, van Cann LM, Kolkman MA, van der Zeijst BA, van Putten JP. Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15C capsular polysaccharides. *Infect Immun*. 2003; 71(11):6192–8. [PubMed: 14573636]
49. Scott JA, Hall AJ, Hannington A, Edwards R, Mwarumba S, Lowe B, Griffiths D, Crook D, Marsh K. Serotype distribution and prevalence of resistance to benzylpenicillin in three representative populations of *Streptococcus pneumoniae* isolates from the coast of Kenya. *Clin Infect Dis*. 1998; 27(6):1442–50. [PubMed: 9868658]
50. Ho PL, Chiu SS, Chan MY, Ang I, Chow KH, Lau YL. Changes in nasopharyngeal carriage and serotype distribution of antibiotic-resistant *Streptococcus pneumoniae* before and after the introduction of 7-valent pneumococcal conjugate vaccine in Hong Kong. *Diagn Microbiol Infect Dis*. 2011; 71(4):327–34. [PubMed: 22000087]
51. McCormick AW, Whitney CG, Farley MM, Lynfield R, Harrison LH, Bennett NM, Schaffner W, Reingold A, Hadler J, Cieslak P, Samore MH, Lipsitch M. Geographic diversity and temporal trends of antimicrobial resistance in *Streptococcus pneumoniae* in the United States. *Nat Med*. 2003; 9(4):424–30. [PubMed: 12627227]
52. Lipsitch M, Colijn C, Cohen T, Hanage WP, Fraser C. No coexistence for free: neutral null models for multistrain pathogens. *Epidemics*. 2009; 1:2–13. [PubMed: 21352747]
53. Hanage WP, Finkelstein JA, Huang SS, Pelton SI, Stevenson AE, Kleinman K, Hinrichsen VL, Fraser C. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. *Epidemics*. 2010; 2(2):80–4. [PubMed: 21031138]

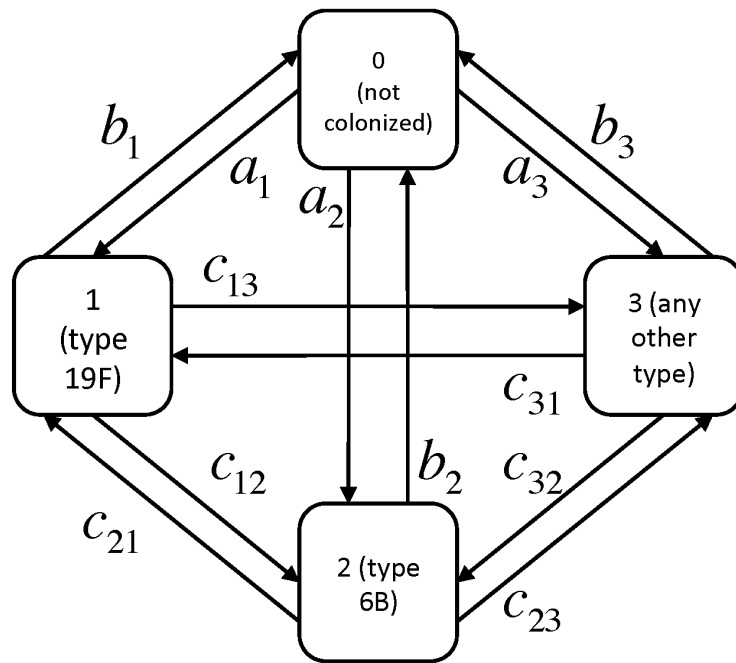


Figure 1. Structure of the Markov transition model for the situation in which two serotypes plus “other” are considered. The model used in the text was similar except that 27 serotypes plus “other” were considered.

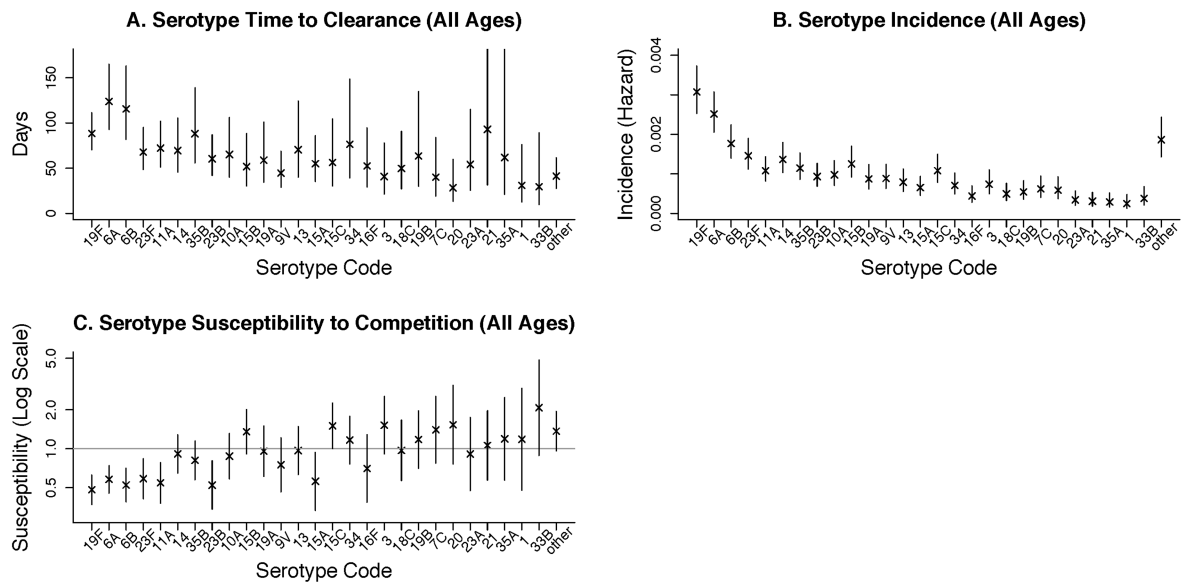


Figure 2. Parameter estimates (with 95% confidence intervals) by serotype, with serotypes ordered from most to least prevalent. A. Mean time to immune clearance b_i . B. Acquisition rate a_i . C. Susceptibility to competition r_i

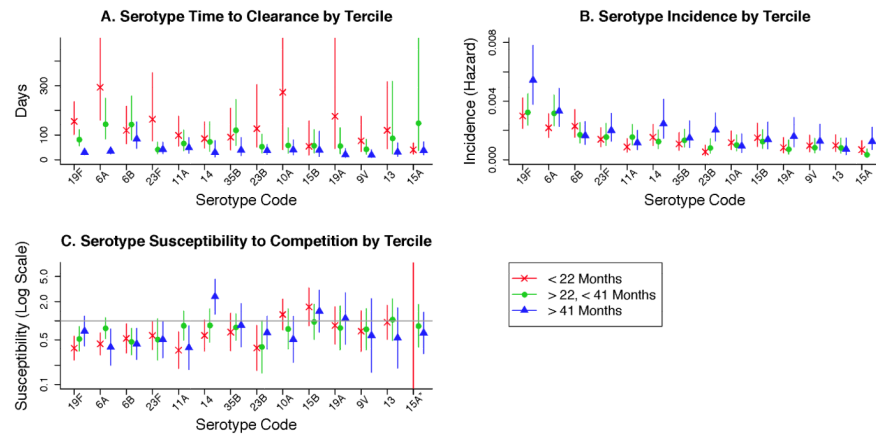


Figure 3.

Parameter estimates (with 95% confidence intervals) by serotype and age tercile, with the top 14 serotypes ordered from most to least prevalent. A. Mean time to immune clearance b_j . B. Acquisition rate a_j . C. Susceptibility to competition r_j

*Susceptibility of 15A at <22 months (not shown) estimated at 0.004741

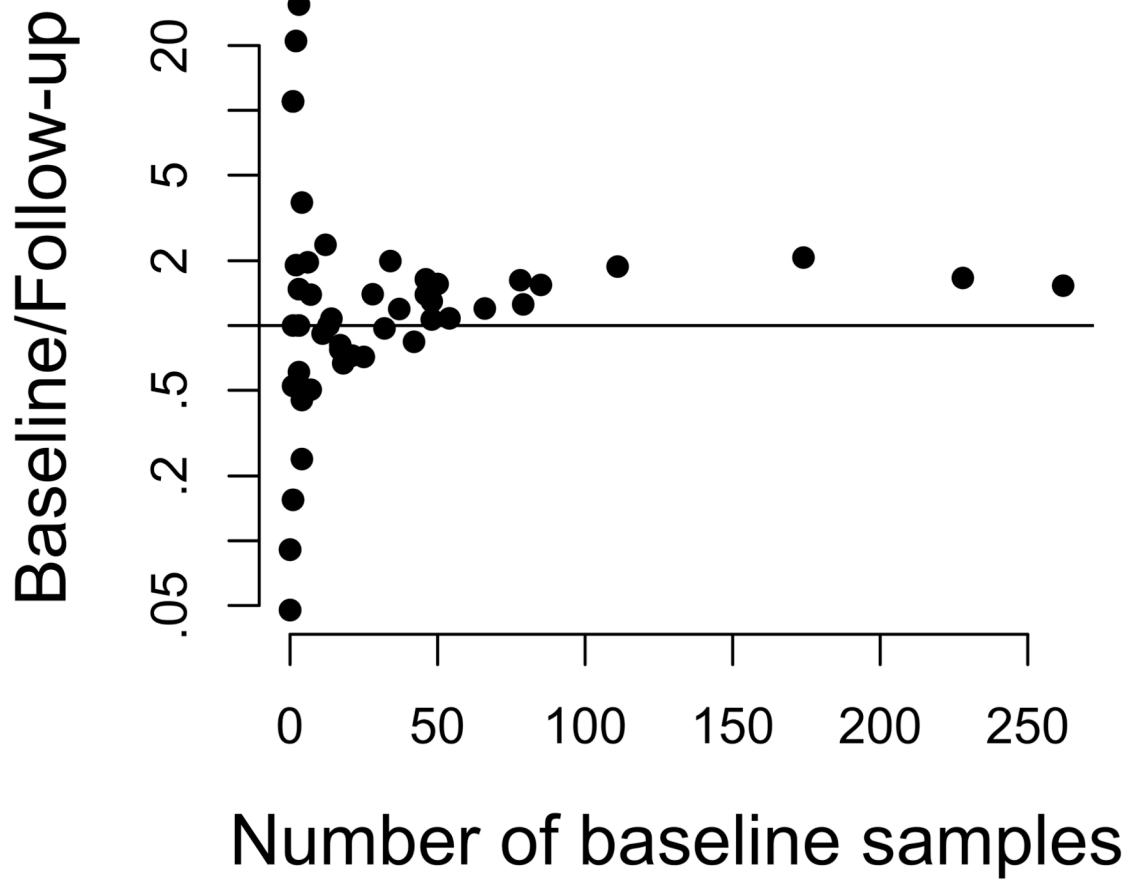


Figure 4.

Number of hosts in whom each serotype was isolated at baseline, compared with the ratio of hosts in whom it was isolated at baseline vs. followup (note log scale on y axis). For rarer serotypes, there was no tendency to be more common at baseline or at followup, whereas more common serotypes tended to be isolated more often at baseline than followup.

Table 1

Model parameters

Parameter	Interpretation
a_i	Acquisition rate for serotype i in persons uncolonized with <i>S. pneumoniae</i>
b_i ($1/b_i$)	Immune clearance rate (time to immune clearance) for serotype i
$c_{ij} = af_i$	Rate of transition from colonization with type i to type j , assumed to be proportional to the acquisition rate for uncolonized persons with type j , times a "susceptibility to competition" r_i specific to the resident serotype i . ^a

^aThis formulation is specific to Model B, which is the model form used for analysis in the paper.

Table 2

Observations used in the analysis, and corresponding indicators and likelihood terms

Observation	Indicator	Likelihood contribution of this observation
Individual not colonized at baseline	$o = 1$	— p_0 , the stationary probability from the Markov model that an individual is in state 0
Individual colonized at baseline with type i	$o = 0$	— p_i , the stationary probability from the Markov model that an individual is in state i
Individual colonized at baseline receives only 1 or 2 swabs, not enough for an observation of any transitions to be made, hence cannot contribute to the likelihood for transitions	$o = 1, y = 0$	none
Individual colonized at baseline with type i is followed from baseline until time u , with at least 3 swabs, and not observed either to clear carriage or to shift to another serotype; i.e. survives all of the hazards	$o = 1, y = 1, q = 1$	$\exp(-h_i u) = \exp[-(b_i + \sum_j c_{ij})u]$
Individual colonized at baseline with type i receives at least 3 swabs, and is observed to clear carriage in the interval (t, u)	$o = 1, y = 1, q = 0, m = 1$	$\frac{b_i}{h_i} [\exp(-h_i t) - \exp(-h_i u)]$, where the bracketed term is the likelihood that some event (where the total hazard is h_i) occurs in the interval (t, u) , and the fraction is the probability that of all possible events, clearance is the first event that occurs
Individual colonized at baseline with type i receives at least 3 swabs, and is observed to switch to type j in the interval (t, u)	$o = 1, y = 1, q = 0, m = 0$	$\frac{c_{ij}}{h_i} [\exp(-h_i t) - \exp(-h_i u)]$, where the bracketed term is the likelihood that some event (where the total hazard is h_i) occurs in the interval (t, u) , and the fraction is the probability that of all possible events, switch to type j is the first event that occurs

TABLE 3

Comparison of models

Model	Expression for c_{ij}	Degrees of freedom	-log likelihood	AIC/2	Comments on model fit
A. Single susceptibility to competition	$c_{ij} = r\alpha_i$	$2n + 1 = 57$	12744.9	12801.9	All estimates stable
B. Multiple susceptibility to competition	$c_{ij} = r_j\alpha_i$	$3n = 84$	12694.0	12778.0	All estimates stable
C. Multiple challenge strength	$c_{ij} = s_j$	$3n = 84$	12726.6	12810.6	Unstable a, s estimates for types 19B, 15C, 7C, 21
D. Full model	$c_{ij} = r_j s_j$	$4n = 112$	12655.5	12767.5	Unstable r, s estimates for all serotypes

Note: An unstable model is defined here as a model with a Hessian matrix evaluated at the numerical optimum which is not negative definite, and thus yields a badly defined asymptotic covariance matrix.

TABLE 4

Serotype correlations (with 95% CI) among epidemiologic estimates and laboratory measurements for 27 serotypes (excluding “other”) except where specified.

	Acquisition (log)	Time to clearance (log)	Susceptibility to competition (log)	Capsule repeat unit carbons* (N=23)	Survival from surface phagocytosis* (N=10)	Time to clearance (log) <22 months	Time to clearance (log) 22-40 months	Time to clearance (log) >=41 months
Prevalence (logit)	0.94 (0.89 to 0.97)	0.71 (0.37 to 0.88)	-0.72 (-0.86 to -0.48)	-0.26 p=0.24	0.72 p=0.02	0.57 (0.11 to 0.83)	0.59 (0.22 to 0.81)	0.14 (-0.34 to 0.57)
Acquisition (log)		0.50 (0.09 to 0.77)	-0.47 (-0.69 to -0.16)	-0.33 p=0.12	0.66 p=0.04	0.47 (-0.02 to 0.77)	0.49 (0.13 to 0.73)	0.06 (-0.40 to 0.50)
Time to clearance (log)			-0.65 (-0.86 to -0.24)	-0.34 p=0.11	0.61 p=0.06	0.70 (0.16 to 0.91)	0.67 (0.07 to 0.92)	0.33 (-0.32 to 0.77)
Susceptibility to competition (log)				0.18 p=0.42	-0.76 p=0.01	-0.43 (-0.74 to 0.05)	-0.45 (-0.76 to -0.01)	-0.16 (-0.62 to 0.39)
Capsule repeat unit carbons*					-0.21 p=0.55	-0.35 p=0.14	-0.12 p=0.61	-0.01 p=0.97
Survival from surface phagocytosis*						0.54 p=0.11	0.49 p=0.15	0.10 p=0.79
Time to clearance (log) <22 months							0.26 (-0.33 to 0.84)	-0.02 (-0.71 to 0.66)
Time to clearance (log) 22-40 months								0.01 (-0.66 to 0.65)

* Spearman correlations.