



Published in final edited form as:

*Epidemiology*. 2011 March ; 22(2): 199–207. doi:10.1097/EDE.0b013e3182087634.

## Prediction of serotypes causing invasive pneumococcal disease in unvaccinated and vaccinated populations

Daniel M. Weinberger<sup>a</sup>, Zitta B. Harboe<sup>b</sup>, Stefan Flasche<sup>c,d</sup>, J. Anthony Scott<sup>e</sup>, and Marc Lipsitch<sup>a,f</sup>

<sup>a</sup>Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA 02115

<sup>b</sup>Department of Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark

<sup>c</sup>Immunisation, Hepatitis and Blood Safety Department, Health Protection Agency, London, UK

<sup>d</sup>Department of Mathematics and Statistics, Strathclyde University, Glasgow, UK

<sup>e</sup>Centre for Geographic Medicine Research-Coast, Kilifi, Kenya

<sup>f</sup>Center for Communicable Disease Dynamics and Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA 02115

### Abstract

**Introduction**—Before the introduction of the heptavalent pneumococcal conjugate vaccine (Prevnar-7), the relative prevalence of serotypes of *Streptococcus pneumoniae* was fairly stable worldwide. We sought to develop a statistical tool to predict the relative frequency of different serotypes among disease isolates in the pre- and post-Prevnar-7 eras using the limited amount of data that is widely available.

**Methods**—We initially used pre-Prevnar-7 carriage prevalence and estimates of invasiveness derived from case-fatality data as predictors for the relative abundance of serotypes causing invasive pneumococcal disease during the pre- and post-Prevnar-7 eras, using negative binomial regression. We fit the model to pre-Prevnar-7 invasive pneumococcal disease data from England and Wales and used these data to (1) evaluate the performance of the model using several datasets and (2) evaluate the utility of the country-specific carriage data. We then fit an alternative model that used polysaccharide structure, a correlate of prevalence that does not require country-specific information and could be useful in determining the post-vaccine population structure, as a predictor.

**Results**—Predictions from the initial model fit data from several pediatric populations in the pre-Prevnar-7 era. Following the introduction of Prevnar-7, the model still had a good negative predictive value, though substantial unexplained variation remained. The alternative model had a good negative predictive value but poor positive predictive value. Both models demonstrate that the pneumococcal population follows a somewhat predictable pattern even after vaccination.

**Conclusions**—This approach provides a preliminary framework to evaluate the potential patterns and impact of serotypes causing invasive pneumococcal disease.

Invasive pneumococcal disease caused by *Streptococcus pneumoniae* (pneumococcus) is a major cause of morbidity and mortality worldwide. There are at least 92 known serotypes,

which differ in their nasopharyngeal carriage prevalence,<sup>1</sup> their tendency to cause invasive disease,<sup>2</sup> and the severity of disease.<sup>3,4</sup>

Until 2000, when the heptavalent pneumococcal conjugate vaccine (Pneumovax-7 or PCV7) was introduced to the US childhood immunization program, the overall rank order of serotypes causing invasive pneumococcal disease was similar in locales worldwide, although the prevalence of specific serotypes varied over time, and some serotypes exhibited distinct geographic patterns.<sup>5-7</sup> Pneumovax-7 targets seven clinically-important serotypes, and it has successfully reduced the burden of pneumococcal disease in areas where it is widely used.<sup>8,9</sup> However, serotypes not covered by the vaccine have filled in the vacated niche, leading to an increase in carriage of these other serotypes.<sup>10</sup> Consequently, non-vaccine serotypes have also increased in disease,<sup>8,9,11,12</sup> although not enough to offset the benefits of vaccination. Given the possibility that serotype replacement could reduce the overall public health benefit of the vaccine, it is important to understand how the bacterial population might change following vaccination.

We use regression models to evaluate which serotypes are most likely to increase in disease following vaccination, and we discuss some potential applications of these approaches.

The incidence rate of invasive pneumococcal disease caused by a given serotype is the product of the prevalence of nasopharyngeal carriage of that serotype and its invasiveness,<sup>13</sup> defined here as the proportion of carriage episodes (or, more accurately, new acquisitions<sup>13</sup>) that result in disease. An ideal model would use invasiveness and carriage prevalence data to predict the disease incidence for each serotype. Because direct estimates of pediatric invasiveness are unstable and available for a relatively small number of serotypes, we first used a known correlate of invasiveness—adult case-fatality ratio (CFR)<sup>14,15</sup>—to generate estimates of invasiveness for 37 serotypes. We combined the regression estimates of invasiveness with carriage prevalence data from England to fit a model to invasive-disease data from England and Wales, and evaluated its performance in various settings. This first model effectively captured the observed serotype distribution in different locations, although there remained substantial unexplained variation in the data. We next fit an alternative model to the pre-vaccine disease data that used estimated invasiveness along with a correlate of carriage prevalence—number of carbons per polysaccharide repeat unit.<sup>16</sup> This second model had a poor positive predictive value, but a good negative predictive value, demonstrating that a single fixed biochemical determinant has an association with the importance of a serotype in invasive disease. While there is still a large amount of unexplained variation in the disease data, the models presented here demonstrate that the distribution of pneumococcal serotypes follows a partially predictable pattern.

## METHODS

### Sources of data

Recent data on pediatric invasive pneumococcal disease cases (less than 5 years old) from the pre- and post-Pneumovax-7 eras were obtained from the Health Protection Agency of England and Wales and the Active Bacterial Core surveillance system of the United States Centers for Disease Control and Prevention (Table 1).<sup>17</sup> For the English data, we focused on the epidemiologic year prior to the introduction of the vaccine (2005/2006) and the most recent epidemiologic year from the post-vaccine era (2008/2009). With the US data, we focused on the 2 years prior to the introduction of Pneumovax-7 (1998-1999) and the most recent three years from which data were available after vaccination (2005-2007). Data for serotypes 6A and 6C were combined together and classified as “6A” for all analyses because not all of the available studies distinguished between them. Pediatric carriage prevalence data from a pre-vaccine English study<sup>18</sup> was used to fit the model. We also sought to

evaluate whether using country-specific carriage data would improve the predictions, so we obtained data both on pediatric carriage and invasive disease of serotypes. We used pre-vaccine carriage and disease data from Kenya (carriage data, Abdullahi et al<sup>19</sup>; disease data, unpublished), and pre- and post- vaccine carriage data from Massachusetts.<sup>10,20</sup> In the carriage datasets from the US and from England and Wales, serotypes 15B and 15C were not distinguished and are known to switch back and forth,<sup>21,22</sup> so we divided the number of 15B/C cases equally between these two serotypes. All carriage prevalence data were natural-log transformed, with serotypes not detected in carriage assigned a value of 0.5 prior to transformation.

We obtained data on adult case-fatality ratio from a large Danish study of adult invasive pneumococcal disease (Brueggemann et al,<sup>3</sup> and unpublished data). The case-fatality ratio was calculated for all serotypes that caused at least 20 cases of invasive pneumococcal disease. Polysaccharide structural data (carbon/repeat unit) was obtained from published reports as described previously.<sup>16</sup> Missing polysaccharide structural data for serotypes 23A and 35F were filled in using the mean values for other serotypes in their respective serogroups.

Invasiveness estimates were obtained from Sleeman et al.<sup>13</sup> and log-transformed prior to use. These invasiveness estimates were considered only for serotypes that had at least 1 carriage and 1 disease isolate detected. Because invasiveness estimates were available for a limited number of serotypes, we then performed linear regression using the case-fatality ratio as a predictor of invasiveness. The results of the regression were then used to generate estimates of invasiveness for all 37 serotypes that had case-fatality data, including serotypes for which there was no direct estimate of invasiveness using the following formula:

$$\text{Estimated invasiveness} = \exp[0.12 * (\text{casefatality } \%) + 5.17]$$

The estimates generated by this regression were then used for all serotypes in downstream analyses, regardless of whether direct estimates were available.

Prediction of the relative abundance of serotypes among invasive pneumococcal disease isolates

**Model 1**—We first fit a log-linked negative binomial regression model to invasive pneumococcal disease count data from England and Wales, using log-carriage prevalence and log-estimated invasiveness as predictors. We considered only the 37 serotypes for which we estimated invasiveness using case-fatality ratios. The resulting model was:

$$\text{Expected \#IPD isolates} = \exp[1.00 * \log(\# \text{carriage isolates}) + 1.08 * \log(\text{estimated invasiveness}) + 2.24]$$

We also used the same model, with the same parameters calculated for England and Wales, to predict serotype frequencies for the United States and Kenya, replacing English carriage data with country-specific carriage data.

**Model 2**—We next fit an alternative model that used estimated invasiveness together with a correlate of carriage prevalence—number of carbons per capsular polysaccharide repeat unit. Again, we used log-linked negative binomial regression and fit the model to pediatric invasive pneumococcal disease data from England and Wales, analyzing the 35 serotypes for which we had data on both estimated invasiveness and polysaccharide structure. The resulting model was:

$$\text{Expected\#IPD isolates} = \exp(0.14 * (\#\text{carbons}/\text{repeat unit}) + 1.16 * \log(\text{estimated invasiveness}) + 3.75)$$

**Evaluation of the fit of the models:** We were primarily interested in the ability of Models 1 and 2 to predict relative contributions of serotypes to invasive disease. Spearman rank correlations were used to assess the relationship between the predicted number of invasive pneumococcal disease cases and the observed. Confidence intervals (CIs) for the correlation coefficients were determined using Fisher's transformation using the `ci2` module in Stata. To further evaluate the ability of the model to predict the relative contribution of serotypes, we calculated the proportion of invasive pneumococcal disease caused by each serotype in a given population, and then generated curves that represented the cumulative proportion of invasive pneumococcal disease for each serotype based on the rank order predicted by the models. Predicted curves with a larger area under the curve (closer to the observed line) represented more accurate predictions.

All statistical analyses were performed using Intercooled Stata, version 9.2 (StataCorp, College Station, TX.).

## RESULTS

### Estimation of pediatric invasiveness using case-fatality ratio

To obtain estimates of pediatric invasiveness for a large number of serotypes, we first performed linear regression between pediatric invasiveness and adult case-fatality ratio. As reported elsewhere,<sup>23</sup> we found a moderate association between these two variables for the 20 serotypes for which we had data (linear regression,  $r^2 = 0.58$ ; eFigure 1, <http://links.lww.com>). We then used the regression coefficients to estimate invasiveness for all 37 serotypes for which we had case-fatality ratio data from a large Danish study.<sup>24</sup> (The estimates for invasiveness are given in eTable 1, <http://links.lww.com>). Importantly, our estimates of invasiveness obtained with this method performed as well or better than the observed invasiveness measures in the downstream models (data not shown).

### Model 1 results for England and Wales pre-vaccination

We next evaluated the ability of carriage prevalence and estimated invasiveness to predict the relative contribution of serotypes to invasive pneumococcal disease among children under the age of 5 years in England and Wales during the pre-Prevnar-7 era (2005-06). Not surprisingly, serotypes with high carriage prevalence and high estimated invasiveness tended to be more common among invasive pneumococcal disease isolates. The coefficients near 1 indicate that the regression recovered the multiplicative relationship between prevalence, invasiveness, and invasive pneumococcal disease incidence. There was a strong correlation between the observed and expected values (Table 2, Figure 1A, Spearman's  $\rho = 0.82$ , (95% confidence interval [CI] = 0.68-0.91)). Additionally, the cumulative proportion curve generated using the predicted serotype ranks closely mirrors the observed cumulative proportion curve (Figure 1B). The predicted values generated from univariate regressions were not as strongly associated with the observed values as were the predicted values from the multivariate regression (carriage prevalence alone ( $\rho = 0.66$ ) or estimated invasiveness alone ( $\rho = 0.24$ )).

### Prediction of serotype ranks in different regions pre- and post-vaccination

We evaluated whether our predictions fit the serotype distributions in different regions in the pre-Prevnar-7 era. We compared the predicted values obtained from the English data against the observed serotype distributions from Kenya and the US. Reflecting the similar serotype

rank orders worldwide, we again found a strong association between the predicted values from the model and the observed number of cases (Table 2).

Because serotype patterns in carriage are similar but not identical among regions, we sought to improve the predictions by calculating the expected invasive pneumococcal disease serotype frequencies using carriage prevalence data from the respective countries. The use of country-specific carriage prevalence data led to a stronger correlation between the observed and expected serotype frequencies in Kenya, but it did not substantially improve the correlation for the US data (Table 2). The predictions made with the US carriage data, however, did fit the US disease data better than the disease data from England and Wales or Kenya.

Next we asked whether these models would have predicted the distribution of serotypes causing invasive pneumococcal disease following the introduction of Prevnar-7. We compared the predicted frequencies calculated from the model using carriage data from England and Wales with the post-vaccine serotype ranks from England and Wales and the US (Table 3, Figures 2A and 2C; non-vaccine serotypes only). There was a reasonably strong association between the predicted and observed serotype ranks, although it was weaker than for the pre-vaccine data. The model had a strong negative predictive value but poor positive predictive value: This is evident when looking at the predicted cumulative proportion curves (Figures 2B and 2D); some of the serotypes predicted to be common contributed little to the cumulative proportion (on the left side of the graph), while the serotypes that were predicted to be rare (right side of the graph) were in fact rare.

Unlike in the pre-Prevnar-7 era, the pre-Prevnar-7 carriage data from Massachusetts did not improve the predictions for the US invasive pneumococcal disease data (Table 3). Serotypes 22F, 19A and 7F, which are among the most important serotypes in the post-vaccine era, were predicted to be ranked 7<sup>th</sup>, 8<sup>th</sup>, and 9<sup>th</sup> out of 30 serotypes. Ranked above these serotypes were types 6A/C, which might cross-react with anti-type 6B antibodies induced by Prevnar-7, and types 1 and 8, which have been shown to exhibit regional variation and temporal fluctuations.<sup>6,7</sup> Also, serotypes 15C, 33F and 38 were predicted to be ranked higher than 7F, 19A and 22F.

### **Model 2: Use of polysaccharide structure rather than prevalence as a predictor**

We next made an alternative model that uses a recently-identified correlate of pre-and post-vaccine carriage prevalence—number of carbons per capsular polysaccharide repeat unit<sup>16</sup>—along with our estimates of invasiveness. In the regression model, polysaccharide structure (number of carbons per polysaccharide repeat) was negatively associated with serotype frequency in disease, while estimated invasiveness was positively correlated with serotype frequency in the pediatric population in England and Wales. The expected values generated from this model were correlated with the observed pre-Prevnar-7 serotype ranks from England and Wales, the US, and Kenya (Table 4). When plotting the cumulative-proportion curves based on the predicted serotype ranks, this model again had a strong negative predictive value (right side of the graph) but relatively poor positive predictive value (left side of the graph) (Figures 3A and 3B). In particular, the model predicted that several serotypes (such as 3 and 9A) would be common that were, in fact, rare among disease isolates. Additionally, serotype 1 was predicted to be the most common serotype; while type 1 is common in England and Wales, it is relatively rare in the US. Comparing the predicted serotype frequencies from Model 1 and Model 2 (eTable 2, <http://links.lww.com>), there is good agreement between the two models for most serotypes. However, Model 2 predicts that types 1, 2, 3, 4, 5, 8, 9A, and 19A will be more common than predicted by Model 1, while Model 1 predicts that serotype 6B would be more common than predicted by

Model 2. In large part, this reflects the fact that polysaccharide structure is a good correlate of carriage prevalence among circulating strains, but this structure does not predict whether a particular serotype will be present in a given population.

Finally, we compared the predictions from this model to the observed serotype ranks in the post-Prevnar-7 era among non-vaccine serotypes. As with the pre-Prevnar-7 population, the model had a good negative predictive value but a poor positive predictive value (Table 4, Figures 3C 3D). Similar to the results from the carriage-prevalence-based model, we could not predict the exact post-Prevnar-7 rank order of serotypes, but the important replacing serotypes (including 7F and 19A) were all ranked in the top half. Additionally, we found that the predictions fit the post-PCV7 England and Wales data substantially better than the post-PCV7 US data. This could be attributed, in part, to serotypes 1 and 8, which are predicted by the model to be common but are rare in the US compared with England and Wales.

## DISCUSSION

Before the introduction of Prevnar-7, the rank order of serotypes causing invasive pneumococcal disease was relatively stable worldwide and over time, but the pneumococcal population has changed considerably now that conjugate vaccines are in widespread use. Serotypes that were less common before the introduction of Prevnar-7 have now increased in carriage frequency and have crept up among invasive pneumococcal disease isolates. It is not yet clear, however, whether disease caused by replacement serotypes will constitute a major concern in the future. In this study, we evaluated whether we could predict the rank order of serotypes among disease isolates before and after the introduction of Prevnar-7 using carriage prevalence data (or a correlate of carriage prevalence) together with estimates of invasiveness. The predictions fit the data from the pre-Prevnar-7 era, and in the post-Prevnar-7 era, the predictions had a good negative but poor positive predictive value.

One potential application of this approach would be to use the estimated invasiveness data to predict how different combinations of serotypes in carriage would affect disease burden. To do this, we used the pre- and post-vaccine carriage data from Massachusetts (2001 vs 2007). The data from these time periods were scaled so that the total numbers of carriage isolates in the two populations were equal, and we used these data in model 1 to predict the disease incidence for each serotype. We predicted that pneumococcal disease in children in the United States would have declined by 28% based solely on differences in invasiveness of the colonizing serotypes. We get similar estimates when we repeat this procedure with the post-vaccine Massachusetts carriage data together<sup>20</sup> with scaled pre-vaccine carriage data from Norway<sup>25</sup> (35% predicted reduction<sup>18</sup>) or with two English carriage studies (28% predicted reduction or 30% predicted reduction using carriage data<sup>26</sup>). These estimates in the reduction of pneumococcal disease are smaller than the 60% reduction in disease among hospitalized cases in the United States,<sup>9</sup> but are consistent with the 40% reduction in the incidence of pneumococcal disease in children in England and Wales.<sup>27</sup> These preliminary findings demonstrate how carriage data, either from population-based studies or vaccine trials, might be used to predict changes in disease burden.

A new 13-valent conjugate vaccine that targets the current Prevnar-7 serotypes as well as serotypes 1, 3, 5, 6A, 7F, and 19A has been licensed in North America and Europe. If effective, this vaccine, PCV13, will eliminate the most common causes of disease and will provide another opportunity to study serotype replacement and to evaluate the predictions of our model. While the incidence of different serotypes will be affected by numerous forces, such as cross-reactivity between serotypes (e.g. anti-serotype 6B antibodies affecting

serotype 6A), the model might help to capture some of the microbiological factors that will influence the post-PCV13 bacterial population structure.

Our model does not account for all variation in the pneumococcal population structure, and this likely reflects the fact that other host and bacterial factors can affect the serotype distribution. Incorporating such factors into future models could improve the positive predictive value of these models. Differences in immunogenicity between serotypes, the immune status of the population, socioeconomic conditions, respiratory co-infections, antibiotic use, and host population structure may all influence the observed patterns of invasive pneumococcal disease incidence. Likewise, bacterial virulence factors other than capsule could affect the ability of a serotype to emerge as a cause of invasive pneumococcal disease. Serotype 1, for example, has a distinct geographic distribution of clones,<sup>28</sup> and recent work has started to explore noncapsular, genetic factors associated with invasiveness.<sup>29,30</sup> Additionally, some serotypes are known to exhibit long-term multi-year fluctuations in incidence.<sup>7</sup> The reasons for such fluctuation are not fully known but could reflect shifts in the genetic background of the dominant serotypes. Because our approach focuses only on the role of capsule—a major determinant of invasiveness—we are not able to predict such fluctuations.

The results of our model further support the notion that the relative invasiveness of serotypes is a fixed property.<sup>2</sup> The model was initially fit using invasiveness and carriage data from England, and we were able to substitute in carriage data from different locations to improve the country-specific fit without changing the invasiveness values.

Following the introduction of Prevnar-7, the serotypes that increased in prevalence were the most common non-vaccine serotypes before the vaccine. As a result, the pre-vaccine disease data is a strong predictor of post-Prevnar-7 serotype ranks. However, while such an approach might provide accurate estimates of post-Prevnar-7 serotype ranks in some situation, it would probably not be effective at predicting the post-PCV13 population structure because the vaccine will eliminate the majority of serotypes causing invasive disease in an unvaccinated population. Our model has the advantage that, in the absence of other information, we can make predictions about which of the remaining serotypes are more or less likely to emerge.

In the pre-Prevnar-7 era, we found that the predicted values based on country-specific carriage prevalence were more tightly correlated with invasive pneumococcal disease rank for that country. However, in the post-Prevnar-7 era, the predicted values calculated with England and Wales pre-Prevnar-7 carriage data were more tightly correlated with the US invasive pneumococcal disease data than were the predicted values calculated with the country-specific prevalence data. One possible reason is that fewer serotypes were detected in the US carriage pre-Prevnar-7 studies than were detected in the England and Wales study. As a result, it is impossible to distinguish between the rare serotypes and to predict which are more likely to increase after vaccination. The pre-vaccine US carriage data were collected during a single season in a single state, so it could be more likely to detect short-term fluctuations in prevalence. Additionally, subjects in the Massachusetts carriage study included only those who visited their pediatricians, and so the serotype distribution in this population might be biased.

Previous work from our group<sup>16</sup> suggests that there could be a biologic explanation for the relative success of different serotypes in nasopharyngeal carriage. The microbial population is not simply a random assemblage of serotypes but is influenced by microbial characteristics that affect interactions with the host. Using the model with polysaccharide

structure as a predictor, a single microbial factor, along with invasiveness, can be used to predict the serotype rank order in several regions.

For the polysaccharide structure model, we found a good negative predictive value for the model but poor positive predictive value. In particular, we would expect that serotypes such as 1, 3, 5, and 9A would be among the most common serotypes in pediatric invasive pneumococcal disease in the US, while in reality they were relatively rare. Types 1 and 5 are common in some parts of the world and were formerly important in the United States, but they are now rare in North America for unknown reasons.<sup>6,31,32</sup> It has been suggested that the difference in relative frequencies of these serotypes between regions is due to difference in blood-culturing practices,<sup>33</sup> although socioeconomic conditions, antibiotic use, differences in the predominant clones, or other factors could also contribute.<sup>31,34</sup>

The models presented here demonstrate that the pneumococcal population structure follows a somewhat predictable pattern, with the second model suggesting that this pattern is influenced by stable microbiologic factors. These approaches also provide a framework for evaluating how serotype composition in carriage could influence the disease burden in the population. The approaches could be used to evaluate potential consequences of future pneumococcal vaccines.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the CDC Streptococcal Laboratory for performing the serotyping of the US invasive pneumococcal disease isolates and the ABCs investigators for providing the invasive pneumococcal disease data from the US, and Robert George and Pauline Kaye for providing the pre- and post-Prevnar-7 dataset on invasive disease for England & Wales, and Moses Ndiritu for providing the Kenya invasive pneumococcal disease data. We thank Rick Malley, Matthew Moore, and Elizabeth Miller for critical reading of the manuscript and discussions and Justin O'Hagan for helpful discussions.

Funding: Supported by the NIH NRSA training program T32 A1007535 and by NIH research grant R01 AI048935 to ML.

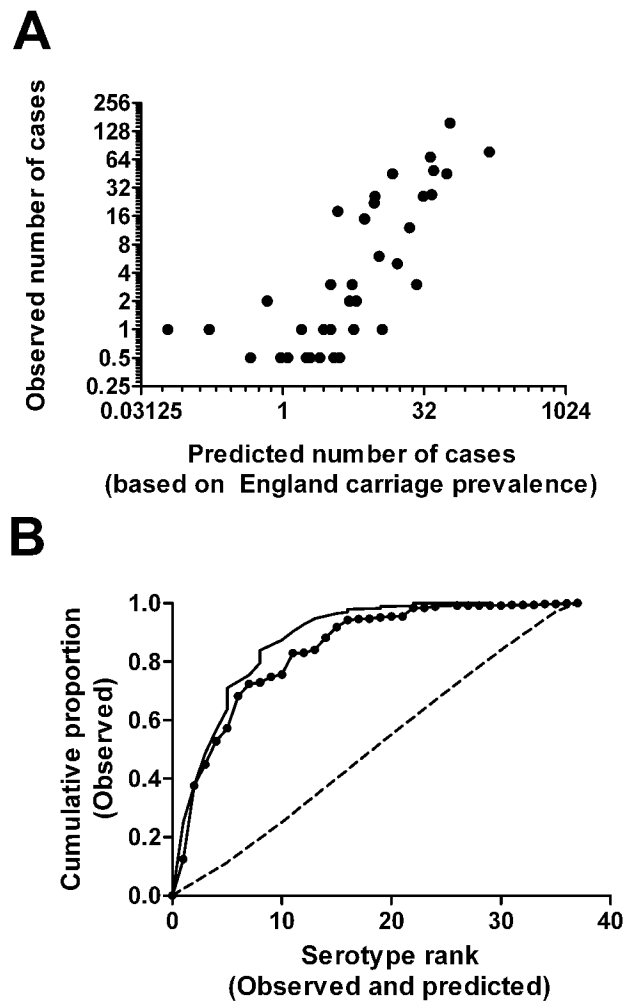
## References

1. Bogaert D, Sluijter M, Toom N, Mitchell T, Goessens W, Clarke S, de Groot R, Hermans P. Dynamics of pneumococcal colonization in healthy Dutch children. *Microbiol.* 2006; 152(2):377–385.
2. Brueggemann AB, Peto TEA, 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):1203–1211. [PubMed: 15346329]
3. Harboe ZB, Thomsen RW, Riis A, Valentiner-Branth P, Christensen J Jr, Lamberts L, Krogfelt KA, Konradsen HB, Benfield TL. Pneumococcal Serotypes and Mortality following Invasive Pneumococcal Disease: A Population-Based Cohort Study. *PLoS Med.* 2009; 6(5):e1000081. [PubMed: 19468297]
4. Weinberger DM, Harboe ZB, Sanders EAM, Ndiritu M, Klugman KP, Rückinger S, Dagan R, Adegbola R, Cutts F, Johnson HL, O'Brien KL, Scott JA, Lipsitch M. Risk of death from pneumococcal pneumonia is a stable serotype-associated property: a meta-analysis. *CID.* 2010 In Press.
5. Hausdorff William P, Bryant J, Paradiso Peter R, Siber George R. Which Pneumococcal Serogroups Cause the Most Invasive Disease: Implications for Conjugate Vaccine Formulation and Use, Part I. *Clin Infect Dis.* 2000; 30(1):100–121. [PubMed: 10619740]



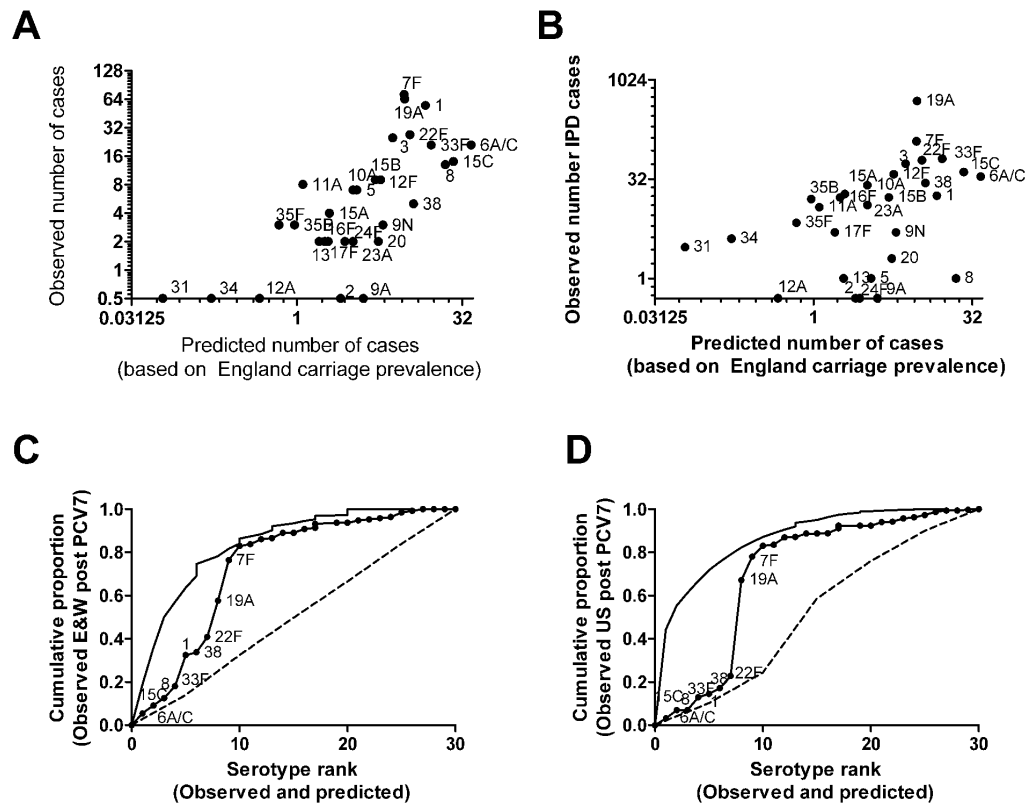
6. Scott J, Hall A, Dagan R, JMS D, SJ E, Fenoll A, Hortal M, Jette L, Jorgensen J, Lamothe F, Latorre C, Macfarlane J, Shlaes D, Smart L, A T. Serogroup-specific epidemiology of *Streptococcus pneumoniae*: associations with age, sex and geography in 7000 episodes of invasive disease. *Clin Infect Dis*. 1996; 22:973–81. [PubMed: 8783696]
7. Harboe ZB, Benfield Thomas L, Valentiner-Branth P, Hjuler T, Lambertsen L, Kaltoft M, Krogfelt K, Slotved Hans C, Christensen Jens J, Konradsen Helle B. Temporal Trends in Invasive Pneumococcal Disease and Pneumococcal Serotypes over 7 Decades. 2010; 50(3):329–337.
8. Kellner James D, Vanderkooi Otto G, MacDonald J, Church Deirdre L, Tyrrell Gregory J, Scheifele David W. Changing Epidemiology of Invasive Pneumococcal Disease in Canada, 1998-2007: Update from the Calgary Area *Streptococcus pneumoniae* Research (CASPER) Study. *Clinical Infectious Diseases*. 2009; 49(2):205–212. [PubMed: 19508165]
9. Pilishvili T, Lexau C, Farley Monica M, Hadler J, Harrison Lee H, Bennett Nancy M, Reingold A, Thomas A, Schaffner W, Craig Allen S, Smith Philip J, Beall Bernard W, Whitney Cynthia G, Moore Matthew R. Sustained Reductions in Invasive Pneumococcal Disease in the Era of Conjugate Vaccine. 2010; 201(1):32–41.
10. Huang SS, Hinrichsen VL, Stevenson AE, Rifas-Shiman SL, Kleinman K, Pelton SI, Lipsitch M, Hanage WP, Lee GM, Finkelstein JA. Continued Impact of Pneumococcal Conjugate Vaccine on Serotypes, Antibiotic Resistance, and Risk Factors for Carriage in Young Children. *Pediatrics*. 2009; 124:e1–11. [PubMed: 19564254]
11. Wenger JD, Zulz T, Bruden D, Singleton R, Bruce MG, Bulkow L, Parks D, Rudolph K, Hurlburt D, Ritter T, Klejka J, Hennessy T. Invasive Pneumococcal Disease in Alaskan Children: Impact of the Seven-Valent Pneumococcal Conjugate Vaccine and the Role of Water Supply. 2010; 29(3): 251–256.10.1097/INF.0b013e3181b1bbed5
12. Isaacman D, McIntosh E, Reinert R. Burden of invasive pneumococcal disease and serotype distribution among *Streptococcus pneumoniae* isolates in young children in Europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines. *International Journal of Infectious Diseases*. 2010; 14(3):197–209.
13. Sleeman KL, Griffiths D, Shackley F, Diggle L, Gupta S, Maiden MC, Moxon ER, Crook DW, Peto TEA. Capsular Serotype-Specific Attack Rates and Duration of Carriage of *Streptococcus pneumoniae* in a Population of Children. *J Infect Dis*. 2006; 194(5):682–688. [PubMed: 16897668]
14. Sjostrom K, Spindler C, Ortvist A, Kalin M, Sandgren A, Kuhlmann-Berenzon S, Henriques-Normark B. Clonal and Capsular Types Decide Whether Pneumococci Will Act as a Primary or Opportunistic Pathogen. *Clin Infect Dis*. 2006; 42(4):451–459. [PubMed: 16421787]
15. Weinberger, DM. Risk of death from pneumococcal pneumonia is a stable serotype-associated property: a meta-analysis. ISPPD-7; Tel-Aviv, Israel: 2010.
16. Weinberger DM, Trzcinski K, Lu Y-J, 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]
17. Hicks Lauri A, Harrison Lee H, Flannery B, Hadler James L, Schaffner W, Craig Allen S, Jackson D, Thomas A, Beall B, Lynfield R, Reingold A, Farley Monica M, Whitney Cynthia G. Incidence of Pneumococcal Disease Due to Non-Pneumococcal Conjugate Vaccine (PCV7) Serotypes in the United States during the Era of Widespread PCV7 Vaccination, 1998-2004. *J Infect Dis*. 2007; 196(9):1346–1354. [PubMed: 17922399]
18. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis*. 2003; 187(9):1424–1432. [PubMed: 12717624]
19. Abdullahi O, Nyiro J, Lewa P, Slack M, Scott JAG. The Descriptive Epidemiology of *Streptococcus pneumoniae* and *Haemophilus influenzae* Nasopharyngeal Carriage in Children and Adults in Kilifi District, Kenya. *Ped Infect Dis J*. 2008; 27(1):59–64.10.1097/INF.0b013e31814da70c
20. Huang SS, Platt R, Rifas-Shiman SL, Pelton SI, Goldmann D, Finkelstein JA. Post-PCV7 Changes in Colonizing Pneumococcal Serotypes in 16 Massachusetts Communities, 2001 and 2004. *Pediatrics*. 2005; 116(3):e408–413. [PubMed: 16140686]

21. Venkateswaran P, Stanton N, Austrian R. Type variation of strains of *Streptococcus pneumoniae* in capsular serogroup 15. *J Infect Dis.* 1983; 147(6):1041–54. [PubMed: 6854063]
22. van Selm S, van Cann LM, Kolkman MAB, van der Zeijst BAM, van Putten JPM. Genetic Basis for the Structural Difference between *Streptococcus pneumoniae* Serotype 15B and 15C Capsular Polysaccharides. 2003; 71:6192–6198.
23. Weinberger Daniel M, Harboe Zitta B, Sanders Elisabeth AM, Ndiritu M, Klugman Keith P, Rückinger S, Dagan R, Adegbola R, Cutts F, Johnson Hope L, O'Brien Katherine L, Anthony Scott J, Lipsitch M. Association of Serotype with Risk of Death Due to Pneumococcal Pneumonia: A Meta-Analysis. *Clinical Infectious Diseases.* 2010; 51(6):692–699. [PubMed: 20715907]
24. Harboe ZB, Thomsen RW, Riis A, Valentiner-Branth P, Christensen JJ, Lambertsen L, Krogfelt KA, Konradsen HB, Benfield TL. Pneumococcal Serotypes and Mortality following Invasive Pneumococcal Disease: A Population-Based Cohort Study. *PLoS Med.* 2009; 6(5):e1000081. [PubMed: 19468297]
25. Vestrheim DF, Hoiby EA, Aaberge IS, Caugant DA. Impact of a Pneumococcal Conjugate Vaccination Program on Carriage among Children in Norway. *Clin Vaccine Immunol.* 2010; 17(3):325–334. [PubMed: 20107006]
26. Hussain M, Melegaro A, Pebody R, George R, Edmunds W, Talukadar R, Martin S, Efstratiou A, Miller E. A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. *Epidemiology and Infection.* 2005; 133(05):891–898. [PubMed: 16181510]
27. Miller, E. Post-PCV serotype changes and variability in observations. ISPPD-7; Tel aviv, Israel: 2010.
28. Brueggemann AB, Spratt BG. Geographic Distribution and Clonal Diversity of *Streptococcus pneumoniae* Serotype 1 Isolates. *J Clin Microbiol.* 2003; 41(11):4966–4970. [PubMed: 14605125]
29. Blomberg C, Dagerhamn J, Dahlberg S, Browall S, Fernebro J, Albiger B, Morfeldt E, Normark S, Henriques-Normark B. Pattern of Accessory Regions and Invasive Disease Potential in *Streptococcus pneumoniae*. *The Journal of Infectious Diseases.* 2009; 199(7):1032–1042. [PubMed: 19203261]
30. Obert C, Sublett J, Kaushal D, Hinojosa E, Barton T, Tuomanen EI, Orihuela CJ. Identification of a Candidate *Streptococcus pneumoniae* Core Genome and Regions of Diversity Correlated with Invasive Pneumococcal Disease. *Infect Immun.* 2006; 74(8):4766–4777. [PubMed: 16861665]
31. Feikin DR, Klugman KP. Historical Changes in Pneumococcal Serogroup Distribution: Implications for the Era of Pneumococcal Conjugate Vaccines. *Clinical Infectious Diseases.* 2002; 35(5):547–555. [PubMed: 12173128]
32. Finland M, Barnes MW. Changes in occurrence of capsular serotypes of *Streptococcus pneumoniae* at Boston City Hospital during selected years between 1935 and 1974. *J Clin Microbiol.* 1977; 5(2):154–166. [PubMed: 14971]
33. Hausdorff WP, Siber G, Paradiso PR. Geographical differences in invasive pneumococcal disease rates and serotype frequency in young children. *The Lancet.* 2001; 357(9260):950–952.
34. Obaro S. Differences in invasive pneumococcal serotypes. *The Lancet.* 2001; 357(9270):1800–1801.



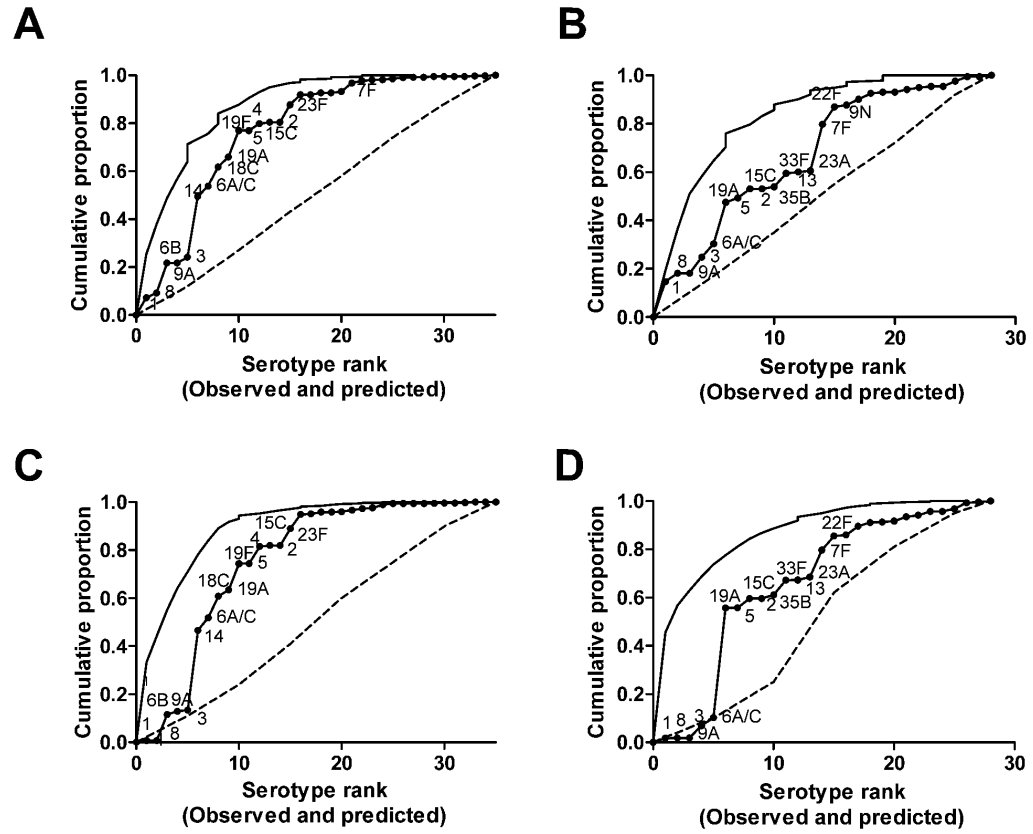
**Figure 1.**

A, Comparison of predicted frequencies with observed serotype frequencies among pediatric (age <5 years) invasive pneumococcal disease isolates during the pre-vaccine era in England and Wales (2005-2006) (Spearman's  $\rho = 0.82$  [95% CI = 0.68-0.91]). B, Cumulative proportion of disease explained by the expected rank order of serotypes (solid line with dots) compared with the actual cumulative proportion curves (solid line) for England and Wales. The dashed line represents the average cumulative proportion of disease explained by a randomly permuted sample. The vertical jumps in the curves are the result of ties in rank.



**Figure 2.**

A, Comparison of observed serotype frequencies among pediatric (age <5 years) invasive pneumococcal disease isolates with predicted frequencies during the post-vaccine era in England and Wales (2008-2009) (Spearman's  $\rho = 0.75$  [95% CI = 0.53-0.87]) and B, the United States (2004-2007) ( $\rho = 0.55$  [95% CI = 0.23-0.76]). C, Cumulative proportion of disease explained by the expected rank order of serotypes (solid line with dots) compared with the actual cumulative proportion curves (solid line) for England and Wales and D, the US. The dashed line represents the average cumulative proportion of disease explained by a randomly permuted sample. The vertical jumps in the curves are the result of ties in rank.



**Figure 3.**

Use of polysaccharide structure, rather than carriage prevalence, to predict serotype ranks pre- and post-Prevna-7. Cumulative proportion of disease explained by the predicted rank order of serotypes (solid line with dots) compared with the actual cumulative proportion curves (solid line) for A, England and Wales, pre-Prevna-7 and C, US, pre-Prevna-7 or in B, England and Wales, post-Prevna-7 or D, the US, post-Prevna-7. The vertical jumps in the curves are the result of ties in rank.

**Table 1**

invasive pneumococcal disease and carriage studies included in analysis

Location	Data type	Data Source	Years analyzed	Age (years)
England and Wales	Disease	Health Protection Agency	2005-2006 (pre-vaccine), 2008-2009 (post-vaccine)	<5
England	Carriage	Brueggemann <sup>18</sup>	1994-2001 (pre-vaccine)	<5
United States	Disease	Active bacterial surveillance (CDC)	1998-1999 (pre-vaccine), 2005-2007 (post-vaccine)	<5
United States (Massachusetts)	Carriage	Huang et al <sup>10,20</sup>	2001 (pre-vaccine), 2007 (post-vaccine)	<7
Kenya (Kilifi)	Disease	Unpublished	1994-2008	<15
Kenya (Kilifi)	Carriage	Abdullahi <sup>19</sup>	2004	<5

**Table 2**

Correlation of observed and predicted serotype ranks; before-Prevnar-7.

Predicted invasive pneumococcal disease ranks based on country-specific carriage data			
	Predicted (England and Wales)	Predicted (USA/MA)	Predicted (Kenya)
Observed invasive pneumococcal disease Ranks			
Observed (England and Wales)	0.82 (0.68-0.91)	0.57 (0.31-0.76)	0.50 (0.21-0.71)
Observed (USA)	0.78 (0.61-0.88)	0.76 (0.56-0.86)	0.61 (0.35-0.78)
Observed (Kenya)	0.51 (0.22-0.72)	0.61 (0.35-0.78)	0.72 (0.52-0.85)

Comparison of observed serotype ranks with serotype ranks predicted using country-specific carriage prevalence data in the model. Spearman's rank correlation coefficients with 95% confidence intervals.

**Table 3**

Correlation of observed and predicted serotype ranks: after-Prevnar-7

	Predicted (England and Wales)	Predicted (USA/MA)
Observed (England and Wales 2008-2009)	0.75 (0.52-0.87)	0.51 (0.18-0.73)
Observed (USA 2005-2007)	0.55 (0.23-0.76)	0.25 (-0.12-0.56)

Comparison of observed serotype ranks with serotype ranks predicted by putting country-specific pre-Prevnar-7 carriage prevalence data into the model. Non-vaccine serotypes only. Spearman's rank correlation coefficients with 95% confidence intervals.



**Table 4**

Model 2: Correlation of observed and predicted serotype ranks using polysaccharide structure as a predictor.

Pre-Prevnar-7	
England and Wales 2005-2006	0.59 (0.31-0.77)
USA 1988-1999	0.59 (0.31-0.77)
Kenya	0.72 (0.50-0.85)
Post-Prevnar-7	
England and Wales 2008-2009	0.44 (0.08-0.70)
USA 2005-2007	0.12 (-0.27-0.47)

Comparison of observed serotype ranks with serotype ranks predicted by the model with polysaccharide structure and CRF. Spearman's rank correlation coefficients with 95% confidence intervals. In the post-vaccine era, only non-vaccine serotypes were included.