Pneumococcal Population Dynamics in the Conjugate Vaccine Era

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:27201746">http://nrs.harvard.edu/urn-3:HUL.InstRepos:27201746</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Pneumococcal Population Dynamics in the Conjugate Vaccine Era

Patrick Mitchell

A Dissertation Submitted to the Faculty of
The Harvard T.H. Chan School of Public Health
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Science
in the Department of Epidemiology
Harvard University
Boston, Massachusetts.
May, 2016
Pneumococcal Population Dynamics in the Conjugate Vaccine Era

Abstract

The introduction of pneumococcal conjugate vaccines (PCVs) in the early years of the 21st century have led to significant changes in pneumococcal epidemiology. Using transmission modeling and genomics based approaches, this dissertation evaluates alterations to the pneumococcal population through the PCV era. Chapter 1 presents a transmission model designed to examine factors that may influence the potential of a previously rare antibiotic resistant lineages to emerge following the introduction of a vaccine targeting more common resistant types, finding that such emergence is more likely in settings with high antibiotic use, high carriage burden, and frequent multiple carriage. Chapter 2 examines the population genomics of pediatric pneumococcal carriage before and after the introduction of PCV-13, finding that the non-vaccine type population composition experienced changes immediately following vaccine introduction but moved back towards its pre-vaccination state over time. Additionally, there is evidence that serotype 3, which is included in PCV-13, has persisted following vaccine introduction, though there are genetic differences between the pre- and post-vaccination population of this serotype. Chapter 3 compares isolates of a single non-vaccine serotype, 33F, collected from carriage and invasive disease, finding evidence that the invasive capacity of this serotype may have declined following the introduction of PCV-13 and that very closely related pairs isolates are disproportionately likely to both be from either carriage or disease. Together, these projects contribute to our understanding of how the pneumococcal population has and will continue to change as PCV use expands.
TABLE OF CONTENTS

List of Figures – iv

List of Tables – vi

Acknowledgements – vii

Introduction – 1

Chapter 1: Carriage Burden, Multiple Colonization, and Antibiotic Pressure Promote Emergence of Resistant Vaccine Escape Pneumococci – 9

Chapter 2: Population Genetics of Pneumococcal Carriage Following PCV-13 Introduction – 38

Chapter 3: Comparison of Carriage and Invasive Serotype 33F Pneumococci Before and After PCV-13 Introduction – 58

Conclusion – 77
LIST OF FIGURES

Figure 1.1: Box diagram of model. Each compartment is further subdivided into vaccinated and unvaccinated populations. The U compartment is uncolonized and subscripts in each colonized (I) compartment indicate the colonizing strains (V – VT-R, N – NVT-S, R – NVT-R). In the vaccinated population, λ_V is replaced by νλ_V. Entry occurs into the S compartment and exit out of the model population occurs at the same rate from all compartments (not shown). All I compartments recover into the U compartment at rate d.

Figure 1.2: Proportion of realizations in which emergence (final NVT-R strain prevalence >10%) was observed by degree of antibiotic pressure. Each point represents a combination of antibiotic pressure (x-axis), carriage burden (lines), and relative acquisition rate k of dual vs. single carriage (panels)

Figure 1.3: Effect of dual carriage on the final prevalence of the NVT-R strain. Each point represents the mean final prevalence of the NVT-R strain at a particular combination of dual carriage relative acquisition rate k (x-axis), carriage burden (lines) and antibiotic pressure (panels).

Figure 2.1: Core genome phylogeny with SCs denoted by color (a). Proportion of population in each sampling period composed of each SC, with shading indicating serotype. Solid colors are PCV-7 type, solid colors with black hatching are PCV-13, and white with colored hatching are not covered by either. Serotype 6A is dotted as it is cross-reactive with 6B, a PCV-7 type, but is itself included in PCV-13.

Figure 2.2: (a) Simpson’s diversity of SCs, excluding the polyphyletic cluster, for each sampling period. (b) Diversity calculated for hypothetical composites of 2007 and 2014 populations, with 2011 diversity shown as dashed line. (c-d) Proportion of population in (c) 2007 and (d) 2014 composed of each SC, ordered by frequency. Gray line is the corresponding distribution from 2011, with dotted lines representing 95% of values from 10000 random samples drawn from the 2011 population.

Figure 2.3: Proportion of the NVT population comprised of each SC. Two additional SCs, SC11 and SC20, had a single NVT isolate and were excluded from this plot.

Figure 2.4: Regression coefficients comparing gene content of the NVT population from 2007 to 2011 and 2014. Black circles correspond to the coefficients with individual genes, with a linear fit to the data shown in black. Fits in which a hypothetical 2014 population was drawn from either the 2007, 2011, or 2014 population are shown in red, blue, and gray, respectively.

Figure 2.5: Proportion of the population in each sampling period comprised of the serotypes included in PCV-13 but not PCV-7

Figure 2.6: Serotype 3 phylogeny, with sampling period shown by color. Isolates collected before the introduction of PCV13, shown in blue, are found primarily on one branch of the tree, while post-introduction isolates, indicated by red, are primarily on the other.

Figure 3.1: Core genome tree of 33F isolates. CCs 717 (orange), 2705 (blue), and 1012 (green) are highlighted. All other leaves belong to CC100. The first bar indicates whether the leaf is a
carriage (blue) or invasive (red) isolate. The second bar indicates the presence of \textit{ermB} and \textit{tetM} (orange) or the \textit{mel-mef} efflux pump (blue).

\textbf{Figure 3.2:} Invasive capacity for the overall 33F population and each clonal complex before (left) and after (right) the introduction of PCV-13, with error bars indicating 95\% confidence bounds. For CCs 717 and 1012, only a post-introduction IC was calculated.

\textbf{Figure 3.3:} Overlapping histograms of the genetic relatedness for pairs of isolates in which one is carriage and the other invasive (between) and where both are from the same source (within), expressed as core genome identity (a) and Jaccard distance of gene content (b). Proportion of isolates from the same source within thresholds of genetic relatedness. Red circles are the result of 100 permutations, with blue lines representing the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentiles of permutation results (c and d).
LIST OF TABLES

Table 1.1 – Description of Compartments
Table 1.2 – Description of Parameters
Table 3.1 – Carriage and Invasive Isolates by Clonal Complex
ACKNOWLEDGEMENTS

First, I would like to thank my research advisor, Bill Hanage, for his guidance and encouragement over the years. As his student, I have been driven to think broadly about my work and to learn from unexpected sources. I would also like to thank Marc Lipsitch for all of the contributions he has made to my education and for lending his knowledge, wisdom, and personality to my research committee. I also thank John Quackenbush for bringing a unique viewpoint to my committee and challenging me to think more deeply about issues I otherwise would have taken for granted.

I would also like to thank the members of the Hanage lab. They have been not just a great resource from which to learn, but also a family with which to laugh (usually) and cry (when necessary). More broadly, I want to thank the CCDD community, past and present, and my friends in the epidemiology department. I feel so fortunate to have been part of such a fun, positive, and supportive group.

Finally, I want to thank the people in my life who helped me grow into the person I am today. I am incredibly privileged to have parents, Lois and Patrick, who have encouraged my educational pursuits throughout my life. I am also eternally grateful to my wife, Alex, for all of her love and support over the years.
INTRODUCTION

The Pneumococcus

_Streptococcus pneumoniae_, or the pneumococcus, is a Gram-positive bacterium that frequently colonizes the human nasopharynx. Carriage is most common in young children, varying regionally with prevalence in children under 5 years of age averaging 65% in low income countries and 48% in lower-middle income countries. In the United States, this figure is around 30%. While most acquisitions of pneumococci never progress beyond asymptomatic carriage, disease can occur when the bacterium invades other body sites. Clinical manifestations include pneumonia, meningitis, bacteremia, and otitis media. Despite the rarity of these outcomes relative to asymptomatic carriage, the high levels of exposure to pneumococcal carriage means the pneumococcus is still responsible for a substantial burden of disease. In 2000, there were an estimated 14.5 million cases of severe pneumococcal disease, resulting in 826,000 deaths, in children under 5 worldwide. In the United States there were approximately 4 million episodes of pneumococcal disease, evenly split between children and adults, in 2004. These episodes led to hospitalization in 445,000 cases and death in 22,000, of which 18,000 were in adults ≥65 years of age.

The polysaccharide capsule, which determines serotype, is the primary virulence determinant of the pneumococcus and the target of all currently licensed pneumococcal vaccines. The heterogeneity of the capsule and its implications for pneumococcal disease was first recognized near the end of the 19th century, and our understanding of the extent of this heterogeneity has grown substantially. Currently, there are more than 90 known pneumococcal serotypes, and this remains the predominant basis on which pneumococci are classified.
Over time, however, our ability to type pneumococci and assign isolates to specific lineages has expanded beyond the capsule. A multi-locus sequence typing (MLST) scheme was developed for the pneumococcus in 1998. This technique groups pneumococci into sequence types (STs) on the basis of the sequence of internal fragments, approximately 450bp in size, from seven housekeeping loci. While isolates of the same ST, meaning they are identical at all seven loci, often share a serotype, the same serotype can be found in distantly related STs and different isolates with the same ST can have different serotypes. The increasing accessibility of whole genome sequence has further increased the resolution with which pneumococci can be classified, using information from the entirety of the ~2.1 million base pairs in the pneumococcal genome, as opposed to the few thousand used for MLST. While there is broad overlap in the categorization produced by MLST and the relationships inferred through whole genome data, sequencing can provide information on within ST variation and allows for the study of the rest of the pneumococcal genetic repertoire.

Each of these typing methods is complicated by recombination. The pneumococcus is naturally transformable, meaning that it is able to take up DNA from the environment and incorporate it into its genome. Due to this process, pneumococci that are distantly related in evolutionary history can share much more recently diverged genetic segments and, correspondingly, very closely related pneumococci can have regions of disproportionately high variation. Population samples show that recombination has frequently affected clinically significant genes, such as those that determine serotype and antibiotic resistance determinants, partially limiting the ability of serotype specific vaccines to control other traits of the pneumococcus.

**Origins of Pneumococcal Vaccination**
Interest in pneumococcal vaccination dates back to the early 20th century. A tetravalent polysaccharide vaccine was tested in 1945, after recognizing that a purified polysaccharide formulation was preferable to earlier attempts at whole cell vaccines. After a brief period of use in the United States in the mid to late 1940s, interest in pneumococcal vaccination waned with the spread of antibiotic usage. However, recognition that antibiotics did not fully defend against pneumococcal mortality, due in part to the rising specter of antibiotic resistant infections, lead to a renewed focus on pneumococcal vaccination. A 14-valent polysaccharide vaccine was licensed for use in the United States in 1977 and was subsequently replaced with a 23-valent formulation, which is still in use, in 1983. While polysaccharide vaccines have been effective in reducing disease caused by the pneumococcal serotypes they target, they do not provoke a strong antibody response in children. As such, they were only recommended for use in children older than 2 years of age with other risk factors and in older populations. A different vaccination strategy was needed to address pneumococcal disease in the broader pediatric population.

This improved strategy came in the form of pneumococcal conjugate vaccines (PCVs). PCVs still target the capsular polysaccharide, but join the polysaccharide with an inert diphtheria toxin variant. Unlike the polysaccharide vaccines, this formulation prompts a T-cell response, leading to protective immunity even in young children. In 2000, a seven valent conjugate vaccine, PCV-7, was recommended for use in all children under the age of two in the United States, ushering in a new era in the control of pneumococci.

The Conjugate Vaccine Era

The past 16 years have seen tremendous progress in the control of pneumococcal disease, due in large part to the use of pneumococcal conjugate vaccines. Overall rates of pneumococcal
disease have dropped substantially, driven by the near elimination of disease caused by vaccine serotypes\textsuperscript{19,20}. Another advantage of PCVs over PPSVs is that they are effective against pneumococcal carriage. As a result, the pediatric population, in which the vaccine was being used, was not the only beneficiary. Herd immunity produced declines in pneumococcal disease amongst adults as well\textsuperscript{19–22}. By targeting serotypes associated with antibiotic resistant lineages, PCV-7 also reduced the rate of resistant pneumococcal infections\textsuperscript{18,23–25}. On the whole, PCV-7 had been a resounding success.

As with invasive disease, the amount of pneumococcal carriage attributable to vaccine types plummeted. In contrast, however, overall carriage rates remained constant\textsuperscript{26,27}. This phenomenon, referred to as serotype replacement, saw pneumococci with capsules not included in the vaccine increase in prevalence and fill the void left by rapidly declining vaccine types. For the most part, these replacement lineages were less troublesome than those they were replacing, as serotypes differ in their propensity to cause disease following carriage acquisition\textsuperscript{28,29}. However, increases in invasive disease and antibiotic resistant infections due to non-vaccine types, especially serotype 19A, began to erode the benefits of PCV-7\textsuperscript{30–32}.

To counter the increasing burden of disease due to replacement serotypes as well as provide protection against serotypes that were responsible for a significant burden of disease in the developing world, extended valence PCVs were developed. A 13-valent formulation, PCV-13, was recommended for use in the United States in 2010\textsuperscript{33}. Thus far, this vaccine has had similar effects to its predecessor, with invasive disease decreasing further and carriage remaining constant despite declines in vaccine types\textsuperscript{3,34}.

As the population composition of \textit{S. pneumoniae} changes in response to PCV-13, there is potential for the emergence of clinically significant non-vaccine type lineages, as occurred
following the introduction of PCV-7. This dissertation takes a multifaceted approach to evaluating alterations to the pneumococcal population through the PCV era. Chapter 1 presents a dynamic transmission model designed to examine factors that may influence the potential of rare lineages to emerge in the wake of vaccination, with specific consideration given to antibiotic resistance. The next chapter takes a broad look at the population genomics of pediatric pneumococcal carriage in Massachusetts; the clonal composition and gene content of the carriage population in the years surrounding PCV-13 introduction, as well as the effect of the vaccine on the new serotypes it covers, are explored in detail. The final chapter focuses on 33F, an increasingly common serotype not included in any current PCV formulation which has been identified as being particularly prone to causing disease. Calculations of the invasive capacity of the serotype before and after PCV-13, as well as a genetic comparison of carriage and invasive isolates, are reported. Together, these projects contribute to our understanding of how the pneumococcal population has already shifted and provide some insight into further changes we might expect in the pneumococcal landscape as PCV use expands.


Chapter 1: Carriage Burden, Multiple Colonization, and Antibiotic Pressure Promote Emergence of Resistant Vaccine Escape Pneumococci

Introduction

*Streptococcus pneumoniae* (the pneumococcus) is a common, widespread bacterial commensal and pathogen. Worldwide, it is estimated to colonize as many as 40 to 60% of young children. (1) While colonization most frequently results in asymptomatic carriage, *S. pneumoniae* is still responsible for a substantial burden of disease. In 2000, there were an estimated 14.5 million episodes of severe pneumococcal disease resulting in 826,000 deaths in children under the age of five. Excluding pneumococcal deaths in HIV-positive children, this accounts for approximately 11% of under-5 mortality. (2)

Pneumococcal strains are classified into more than 90 different serotypes based on the antibody reactivity and chemical structure of the polysaccharide capsule, the major surface antigen of pneumococci. Pneumococcal conjugate vaccines (PCVs) have proved effective in reducing the populations of pneumococci bearing vaccine type (VT) capsules, but only target a handful of serotypes. PCV use has been followed by the emergence of strains of non-vaccine serotype (NVT) in carriage (3-4) and, to a lesser degree, disease. (5) Among these are some VT lineages which have evaded the vaccine by acquiring an NVT capsule by recombination, a frequent process in pneumococci. (6) Such ‘capsule switch’ variants have successfully spread across the United States and internationally following the adoption of the heptavalent PCV (PCV-7) in 2000. (6-7) In addition to genes involved in capsule biosynthesis, other genes primarily found in VT lineages have the potential to enter NVT lineages through recombination. (8) In the absence of vaccination, these genes may be rare or absent in NVT lineages. Whether they first appear in NVT lineages before or after vaccine introduction, these genes may help the NVT lineages in which they are found expand into niches cleared of VT lineages. Genes that
affect antigenicity, pathogenicity, and antibiotic resistance are of particular concern given the relevance of these phenotypes to human disease. These genes have also shown a high frequency of observed recombination, reflecting one or both of two factors: strong selective pressure favoring recombinants and/or intrinsically higher rates of genetic recombination. (6, 9)

The emergence of new vaccine escape variants, particularly those exhibiting high-level antibiotic resistance, will continue to cause concern as PCV use expands. According to the March 2014 Vaccine Information Management System (VIMS) Global Vaccine Introduction Report, (10) 39 countries currently have plans to introduce a PCV. Of the countries planning introduction, 22 are classified as middle income and 11 as low income. The epidemiology of *S. pneumoniae* in such settings is highly variable and can be quite different from that in higher income settings. (11-12) Additionally, antibiotic usage can differ sharply between countries. (13) As such, the effect of PCV introduction on the emergence of resistant vaccine escape variants may differ from what has been seen thus far.

Of the major lineages that emerged following the introduction of PCV-7, most appear to have been present at a low frequency prior to vaccination. (6, 8, 14) As such, we sought to evaluate factors that may influence the emergence of a rare antibiotic-resistant pneumococcal lineage following the introduction of vaccination targeting more common resistant types. Here, we employ a stochastic dynamic model of pneumococcal carriage to assess the role of antibiotic pressure, carriage burden, and multiple colonization in promoting this emergence.

**Model Structure**

Acquisition and clearance of *S. pneumoniae* carriage is simulated here using a three-strain SIS model designed to avoid implicit promotion of coexistence by limiting competition
between strains. (15) A condensed model schematic is shown in Figure 1.1. The model uses an open population in which people exit the population from each compartment at rate $u$, here set at $1/60$ month$^{-1}$. People enter the population at the same rate multiplied by the initial population size in order to produce a roughly stable population size covering a 5 year age range.

**Figure 1.1 -** Box diagram of model. Each compartment is further subdivided into vaccinated and unvaccinated populations. The U compartment is uncolonized and subscripts in each colonized (I) compartment indicate the colonizing strains (V – VT-R, N – NVT-S, R – NVT-R). In the vaccinated population, $\lambda_V$ is replaced by $v\lambda_V$. Entry occurs into the S compartment and exit out of the model population occurs at the same rate from all compartments (not shown). All I compartments recover into the U compartment at rate $d$.

Individuals may be colonized either by a single strain or any combination of two strains. Full details of each compartment are given in Table 1.1 and parameters are described in Table 1.2. Dual colonization with a single strain is treated as being identical to single colonization with
that strain. Those currently colonized have their susceptibility to acquiring another strain reduced by a relative susceptibility factor $k$. If an individual colonized by two strains acquires a third, it is equally probable that either of the two original colonizing strains will be replaced by the new one. An important feature of the model structure is that individuals in the dual-colonized compartment can acquire new colonization with one of the two strains they are already carrying. If the new colonizing strain is the same as that which it is replacing, the individual remains in the same compartment, colonized by the same two strains; if it replaces the other, such that the host is now ‘dual-colonized’ with the same strain, the host is moved into the appropriate single carriage compartment. Clearance of colonization occurs at the same rate $d$ from all colonized compartments. Each compartment in the model is duplicated to distinguish between vaccinated and unvaccinated individuals. A fraction $f$ of new entrants, here set to 90%, are vaccinated at birth and thus become part of the vaccinated population, while the rest enter the fully susceptible population.

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Unvaccinated</th>
<th>Vaccinated</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>VS</td>
<td>VV</td>
<td>Uncolonized</td>
</tr>
<tr>
<td>I</td>
<td>NS</td>
<td>NV</td>
<td>Colonized with VT-R strain only</td>
</tr>
<tr>
<td>I</td>
<td>RS</td>
<td>RV</td>
<td>Colonized with NVT-R strain only</td>
</tr>
<tr>
<td>I VS</td>
<td>VNS</td>
<td>VNV</td>
<td>Colonized with both VT-R and NVT-S strains</td>
</tr>
<tr>
<td>I NS</td>
<td>I VRS</td>
<td>I VRV</td>
<td>Colonized with both VT-R and NVT-R strains</td>
</tr>
<tr>
<td>I RS</td>
<td>I NRS</td>
<td>I NRV</td>
<td>Colonized with both NVT-S and NVT-R strains</td>
</tr>
</tbody>
</table>
Table 1.2 – Description of Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>1.5-2.5 ( \times 10^{-5} )</td>
<td>Transmission parameter - corresponds to ( R_0=1.5-2.5 ), w. 1 month duration and total population=( 10^5 ) (11-12, 30)</td>
</tr>
<tr>
<td>( d )</td>
<td>1 month(^{-1} )</td>
<td>Rate of clearance, corresponds to 1 month duration of carriage</td>
</tr>
<tr>
<td>( u )</td>
<td>1/60 month(^{-1} )</td>
<td>Entry/exit rate</td>
</tr>
<tr>
<td>( a )</td>
<td>1.0-1.05</td>
<td>Transmission advantage of vaccine type and recombinant strains</td>
</tr>
<tr>
<td>( k )</td>
<td>0-0.5</td>
<td>Relative susceptibility of those already colonized (0=no dual colonization) (21-25)</td>
</tr>
<tr>
<td>( v )</td>
<td>0.5</td>
<td>Reduction of transmissibility of vaccine type strain to vaccinated individuals (16)</td>
</tr>
<tr>
<td>( f )</td>
<td>0.9</td>
<td>Fraction of new entrants into the population that are vaccinated</td>
</tr>
</tbody>
</table>

\[
\lambda_N = \beta(I_{NS}+I_{NV}+0.5(I_{VNS}+I_{VNV}+I_{VRS}+I_{VRV}))
\]

\[
\lambda_N = \beta(I_{NS}+I_{NV}+0.5(I_{VNS}+I_{VNV}+I_{VRS}+I_{VRV}))
\]

\[
\lambda_R = a\beta(I_{RS}+I_{RV}+0.5(I_{VRS}+I_{VRV}+I_{NRS}+I_{NRV}))
\]

Each strain is defined by a combination of two characteristics. The first represents serotype, dichotomized into VT and NVT. Vaccinated individuals are protected from colonization with a VT strain by a factor \( v \), set to produce a 50% reduction in susceptibility. (16) The other characteristic is a selective advantage conferred by antibiotic resistance implemented as a multiplicative increase \( a \) in the transmission parameter. This is meant to correspond to an
increased availability of hosts to resistant strains, as a certain proportion of the population may be taking an antibiotic at any given time. We conceptualize this resistance trait as being conferred through large scale genomic changes, such as mosaic penicillin binding proteins conferring β-lactam resistance or mobile elements introducing macrolide resistance genes, (17) rather than being introduced by a single point mutation. As such, *de novo* generation of resistance is not included in this model.

Three combinations of these two traits are considered in the model. The first is an antibiotic-resistant VT strain (VT-R) and the second is an antibiotic-susceptible NVT strain (NVT-S). The former has a potential advantage in unvaccinated populations as a result of being antibiotic-resistant, while the latter is at an advantage in vaccinated populations due to its NVT capsule. The third strain combines the advantageous traits of the first two, being both antibiotic resistant and having an NVT capsule. The fourth possible combination, in which the strain would be inhibited by both vaccination and antibiotic use, is not considered as it would be quickly eliminated in the scenarios considered here.

The initial pneumococcal population is composed primarily of the VT-R and NVT-S strains in equal proportion. This is meant to reflect a hypothetical situation in which the vaccine being introduced targets the serotypes currently associated with the great majority of antibiotic resistance, similar to the situation when PCV7 was first introduced in the US. (18) The starting prevalence of the NVT-R strain is set to 1% of the combined initial prevalence of the VT-R and NVT-S strains. This starting prevalence was chosen based on the well-documented emergence of serotype 19A ST320 pneumococci following the introduction of PCV7. In a study of the response of the pneumococcal population to vaccination, carriage isolates were collected from children attending Massachusetts primary care practices in 2001 and 2004. In the combined
sample size of over 300, a single isolate was observed with the combination of serotype (19A) and sequence type (ST 320) that went on to become successful. (19-20) A true prevalence of 1% is near the upper bound of the prevalence in the underlying population that could be expected to produce this observation under the binomial distribution.

**Model Implementation**

Using this model, we sought to compare the potential for emergence across a range of settings. The initial combined carriage prevalence of the VT-R and NVT-S strains was set to either 30%, 50%, or 60% with corresponding baseline $R_0$ values of approximately 1.5, 2.0, and 2.5 in order to simulate low, moderate, and high burden settings. (11-12) The resistance advantage for the VT-R and NVT-R strains varied from 1.0 (selective neutrality) to 1.05 (5% advantage) to simulate increasing levels of antibiotic pressure.

A proportion of carriers are known to be colonized with multiple serotypes, suggestive of multiple acquisitions. Colonization with two (or, in rare cases, three) serotypes has been reported at levels consistent with those colonized having experienced an approximately 80% reduction in susceptibility to colonization with a new serotype, corresponding to a relative susceptibility of 20% when compared with uncolonized hosts. (21-22) Other studies have attempted to estimate the degree of competition in the acquisition of a new strain and found the relative susceptibility of those already colonized to range from <10% to nearly 50%. (23-25) To examine the impact of multiple carriage, the relative susceptibility of those currently colonized compared to those not currently colonized ranged from 0 (competitive exclusion with no dual colonization) to 50%.

Aside from this barrier to dual colonization, this model does not include any within-host competition and both strains in a dual colonized host are assumed to be equally likely to
transmit. An alternative scenario in which the two resistant strains are given a lower relative fitness as compared to the sensitive strain, resulting in the resistant strains being less likely to transmit from a host also colonized with the sensitive strain, is considered in the supplementary material.

The model was run as a series of stochastic simulations using the adaptive tau-leaping method implemented in R through the “adaptivetau” package. (26) For all parameter sets, 50 realizations were conducted. The initial conditions of each simulation were meant to approximate pre-vaccine equilibrium conditions and the immediate introduction of vaccination. The combined starting prevalence of the VT-R and NVT-S strains was set to roughly the equilibrium prevalence predicted by the $R_0$ used and the starting vaccine coverage was 90%, equal to the proportion of the incoming population vaccinated. The effect of lower or no vaccine coverage is evaluated in the supplementary material [Figures S1-2]. Each simulation covered a 10 year timespan and had a starting population of 100,000 plus the starting number with the new strain. The final prevalence of the NVT-R strain was taken as the primary outcome and emergence was defined as a final NVT-R strain prevalence greater than 10%. Outcomes for the VT-R and NVT-S strains are detailed in the supplementary material [Figure S3].

Results

Effect of Antibiotic Pressure

In order to assess the potential for emergence of a new resistant strain over a range of antibiotic pressures, simulations were conducted with the advantage parameter for the VT-R and NVT-R strains set between 1.0 and 1.05 in increments of 0.01. Varying the transmission advantage of the antibiotic resistant strains had predictable effect on emergence. At lower values,
the NVT-R strain was unable to overcome the greater initial prevalence of the NVT-S strain and could not colonize a substantial portion of the population within 10 years. Under neutral conditions no emergence was observed. With a 1% advantage, emergence was observed in a minority of realizations for the high-burden setting. At 3% and above, emergence was nearly universal, indicating that antibiotic use is a key determinant of the emergence of resistant lineages following PCV introduction. When the selective advantage due to antibiotic resistance was 2%, the potential for emergence varied by carriage burden and relative susceptibility to dual carriage [Figure 1.2].
Figure 1.2 – Proportion of realizations in which emergence (final NVT-R strain prevalence >10%) was observed by degree of antibiotic pressure. Each point represents a combination of antibiotic pressure (x-axis), carriage burden (lines), and relative acquisition rate $k$ of dual vs. single carriage (panels).
Greater Emergence at Higher Carriage Burden

As plans progress to introduce PCVs into new settings, it is worthwhile to consider the prospects for emergence across a range of carriage burdens reflecting the epidemiological situation in contexts other than the US and Europe. To this end, the pneumococcal carriage prevalence was set to ~30, 50, or 60% through a combination of the starting prevalence and transmission parameter. As noted above, across carriage burdens, emergence occurred in all realizations when the selective advantage due to antibiotic resistance was 4% or greater, and emergence was never observed when there was no resistance advantage. At intermediate levels of antibiotic pressure, however, emergence was more likely in higher burden settings [Figure 1.2]. The NVT-R strain was also responsible for a higher proportion of colonizations in the higher burden settings [Figure S4]. While the new strain was never completely absent by the end of the simulation, indicating that important changes are possible over a longer time frame than we consider here, these results suggest that emergence of new antibiotic resistant lineages may occur more rapidly and with less antibiotic pressure in settings with higher pneumococcal burdens. When the relative fitness of resistant strains was low as compared to sensitive strains, however, this effect may be reversed [Figure S5].

Dual Carriage Promotes Emergence

We evaluated the effect of competition and dual carriage by altering the relative susceptibility of those currently colonized compared to those not currently colonized between 0 (no dual carriage) and 0.5 in increments of 0.1. In general, increases in dual carriage allowed for greater spread of the NVT-R strain within the population. Across all scenarios considered, the
mean final prevalence of the NVT-R strain was strongly correlated with relative susceptibility to dual carriage, though the effect was most pronounced at intermediate values for the resistance advantage [Figure 1.3]. When resistance conferred a fitness cost, however, increased dual carriage inhibited emergence of the NVT-R strain [Figure S5]. Similar to the effect of carriage burden, the role of dual carriage was of secondary importance at high or low antibiotic pressures, as the new strain either approaches its maximum prevalence or is present at only very low levels, respectively.
Figure 1.3 - Effect of dual carriage on the final prevalence of the NVT-R strain. Each point represents the mean final prevalence of the NVT-R strain at a particular combination of dual carriage relative acquisition rate $k$ (x-axis), carriage burden (lines) and antibiotic pressure (panels).

Discussion
As PCVs are introduced into more immunization programs, there will be continued alterations of the pneumococcal population. Here we evaluated several factors that may influence the emergence of previously unseen antibiotic resistant lineages following the widespread introduction of pneumococcal vaccination. As would be expected, the degree of antibiotic pressure in the population is highly influential and appears to guarantee the emergence of new resistant types if sufficiently strong. When selection for antibiotic resistance was not as strong, increases in carriage burden and allowance for dual colonization also promoted emergence. These factors can both be seen as increasing the availability of new hosts, which are better exploited by the new resistant strain.

The observed effect that dual colonization promotes emergence suggests that intra-specific and intra-host competition may be important factors in determining the success of new pneumococcal variants. As this model simply dichotomizes pneumococci into vaccine type and non-vaccine type and doesn’t consider any immunological factors beyond vaccination, it cannot fully capture these dynamics. Significant work has gone into modeling this type of competition in more detail using individual based models. (27-28) While adding more realism in this respect would structurally alter the model, the qualitative findings would likely be similar, as the degree of antibiotic pressure was a much stronger driver of emergence. Additionally, previous models with differing structures have predicted expansion of rare NVT serotypes following vaccination yet did not consider resistance, (29) and in another case the maintenance of penicillin resistance even after vaccination targeting all currently resistant serotypes in a setting with significant selection for resistance in a model that did not account for multiple carriage. (30) We have integrated these factors into a single model with the conclusion that antibiotic pressure is likely the most important driver of the emergence of resistant, NVT lineages.
Our findings also align with observed changes in the distribution of antibiotic resistance following the introduction of PCV-7. Non-susceptibility to multiple classes of antibiotics rose in NVT pneumococci as they replaced vaccine types (31) and common lineages were displaced by more resistant clones within specific NVT serotypes. Examples of this phenomenon include sequence type (ST)320 replacing ST199 in serotype 19A (32) and CC176 being supplanted by ST386 in serotype 6C. (33) This combination of observed data and other modeling work lend credence to our finding that the expansion of a previously rare antibiotic resistant lineage is probable following the introduction of a PCV targeting common resistant types. So long as vaccination targets only a subset of pneumococcal serotypes, antibiotic pressure will likely lead to the emergence of resistant lineages. This may be facilitated in settings that combine high or intermediate drug pressure with high carriage prevalence.

Despite this, conjugate vaccines have been and will continue to be a valuable tool for reducing the public health burden imposed by S. pneumoniae. Substantial reductions in pneumococcal disease, particularly due to pneumococci non-susceptible to penicillin and to multiple antibiotics, were observed in the aftermath of PCV-7. (34) This has occurred despite carriage prevalence remaining roughly constant, (3-4) likely due to differences in invasive capacity between serotypes. (35) The strategy of vaccinating against the serotypes associated with the most worrisome pneumococcal lineages has indeed proven successful, but the plasticity of the pneumococcus necessitates continued vigilance to monitor for emerging variants. (7-8, 14, 31-33)
References


Supplementary Material

Effect of Lower Vaccine Coverage

In order to evaluate the extent to which our results would be altered by lower vaccine coverage, we ran five realizations with vaccine coverage reduced from 90% to 50% at each combination of carriage burden, antibiotic pressure, and dual carriage level. The mean final prevalence of the NVT-R strain was highly similar between simulations with 50 and 90% vaccine coverage for all parameter sets, indicating that high vaccine coverage is not necessary to reproduce the observed emergence pattern [Figure S1]. Vaccine coverage of 50% was also sufficient to eliminate the VT-R strain within 10 years in all simulations. With no vaccination, the NVT-R strain was never able to infect greater than 10% of the population. The VT-R and NVT-S strains infected similar proportions of the population with no vaccination and no resistance advantage, but as antibiotic pressure increased, the VT-R strain became dominant [Figure S2].

![NVT-R Mean Final Prevalence](image)

**Figure S1**: Comparison of final prevalence of NVT-R strain with 50% vaccine coverage to results from the main model with 90% coverage. Each point represents the mean final prevalence of the NVT-R strain at some combination of carriage burden, antibiotic pressure, and dual carriage level over 5 realizations and 50% coverage and 50 realizations at 90% coverage. The red line represents a perfect 1:1 correspondence between the mean final prevalence of the NVT-R strain between the two levels of vaccine coverage.
**Figure S2:** Ratio of the mean final prevalence of the VT-R and NVT-S strains across levels of antibiotic pressure in the absence of vaccination. Each point represents the mean outcome over five realizations at each combination of carriage burden, antibiotic pressure, and allowance for dual carriage.
Relationship between Final Prevalences of Strains

An additional 5 realizations of the model with 90% vaccine coverage were run at each parameter set to track the mean final prevalence of each of the three strains as well as the proportion of the population uncolonized. The VT-R strain was completely eliminated from the population in all simulations. There was an inverse relationship between the mean final prevalence of the NVT-S and NVT-R strains, with the overall final prevalence remaining roughly constant across parameter sets [Figure S3]. This indicates that emergence of the NVT-R strain comes at the expense of the NVT-S strain, rather than increasing the overall carriage burden.

Figure S3: Total final prevalence and final prevalence of NVT-S strain in relation to final prevalence of NVT-R strain. Each point represents the mean results over five realizations at each combination of burden, antibiotic pressure, and dual carriage. The final prevalence of the NVT-S strain (squares) decreases as mean final prevalence of the NVT-R strain increases, while the total final prevalence (triangles) remains roughly constant.
Effect of Carriage Burden on NVT-R Proportion

To evaluate whether the effect of higher carriage burden increasing emergence was solely due to a higher equilibrium prevalence, we compared the proportion of colonized carrying the NVT-R strain between carriage burdens. Using the same set of 5 realizations for each parameter set, the mean final prevalence of the NVT-R strain was divided by the overall mean final prevalence of colonization. At intermediate values of antibiotic pressure, the NVT-R strain was responsible for the greatest proportion of colonizations in the high burden setting and the least in the low burden setting [Figure S4].

Figure S4: Proportion of the overall final prevalence due to the NVT-R strain. Each point represents the mean final prevalence of the NVT-R strain divided by the overall mean final prevalence over five realizations at each combination of burden, antibiotic pressure, and dual carriage.

Decreased Relative Fitness Inhibits Emergence
Genetic changes encoding antibiotic resistance are typically thought to impose a fitness cost in antibiotic free environments. In order to explore how such a cost could affect the emergence of the NVT-R strain, we altered our model such that hosts colonized with both a susceptible and resistant strain could be more likely to transmit the susceptible strain than the resistant one. The relative fitness of the resistant strain was altered between 0.5 (resistant half as likely to transmit as sensitive) to 1.0 (equal probabilities, as in main model) in increments of 0.1. For each value for relative fitness, 5 realizations were conducted at each combination of carriage burden, antibiotic pressure, and dual carriage evaluated in the main model. As would be expected, lower relative fitness strongly inhibited the emergence of the NVT-R strain. This effect became more pronounced as the allowance for dual carriage increased and was less pronounced in the low carriage burden setting [Figure S5].
**Figure S5**: Mean final prevalence of NVT-R strain across values for relative fitness of resistant strains, with relative fitness defined as the ratio of transmission of a resistant strain versus a sensitive strain from a dual colonized host. Each panel contains the results for each of the three carriage burdens for a specific combination of antibiotic pressure (a) and dual carriage allowance (k).
Chapter 2: Population Genetics of Pneumococcal Carriage Following PCV-13 Introduction

INTRODUCTION

*Streptococcus pneumoniae* is a common bacterial colonizer of the human nasopharynx, particularly among children\(^1\). In Massachusetts, it has consistently been found in approximately 30% of children under the age of 7 between 2001 and 2011\(^2\). While colonization rarely progresses beyond asymptomatic carriage, the ubiquity of the pneumococcus leads to a substantial burden of disease, causing an estimated 4 million disease episodes, including 445,000 hospitalizations and 22,000 deaths in the United States in 2004\(^3\).

Conjugate vaccination has been a major advance in the reducing pneumococcal disease. The seven-valent pneumococcal conjugate vaccine (PCV-7), introduced in the United States in 2000 was highly effective in reducing overall rates of pneumococcal disease, as vaccine type (VT) pneumococci were responsible for the vast majority of cases\(^4\)–\(^6\). Carriage of vaccine serotypes also declined, though overall carriage prevalence remained roughly constant due to serotype replacement\(^2\),\(^7\),\(^8\).

Despite lower overall rates of pneumococcal disease, increases were seen in the incidence of disease due to the replacement non-vaccine type (NVT) population. Serotype 19A in particular became a significant cause of invasive disease\(^5\),\(^9\),\(^10\). The thirteen-valent vaccine (PCV-13), introduced in 2010, extended coverage to six additional serotypes, including 19A, beyond those included in PCV-7, and has resulted in further reductions in pneumococcal disease\(^11\). As with PCV-7, however, overall carriage prevalence has not changed substantially\(^2\). Concerningly, serotype 3, a highly invasive serotype included in PCV-13 appears not to have declined as much as the other newly added serotypes\(^2\),\(^11\)–\(^14\). Given the potential for disease to arise both from
replacement NVTs and persistent VTs, it remains important to monitor changes to the pneumococcal carriage population.

Pediatric pneumococcal carriage in Massachusetts has been extensively studied since shortly after the introduction of PCV-7. The effects of vaccination can be seen both in the prevalence of specific lineages as well as in broader population metrics. The apparent effects of vaccination are variable depending on how the population is characterized and the timescale over which it is examined. Serotype diversity was found to have increased then stabilized following the introduction of PCV-7\textsuperscript{15}, reflecting the selective impact of vaccines and the period while carriage replacement was taking place. Interestingly, minimal changes were found when comparing the presence and absence of specific pneumococcal genes in this population between 2001 and 2007, suggesting that the overall genetic composition of the population was not much changed other than a few loci found only in one of the loci conferring vaccine serotype 6B\textsuperscript{16}. Another study considering MLST profiles found no significant change in diversity or population composition in the immediate aftermath of PCV-13\textsuperscript{17}. With more time elapsed since PCV-13 introduction, it is possible to evaluate the longer term effects of this vaccine.

Here we examine population–scale genetic changes in carriage pneumococci amongst children in Massachusetts since the introduction of PCV-13. Using genomic sequencing data for isolates collected between 2000 and 2014, we analyze alterations to the clonal composition, defined on the basis of core genome variability, and gene content of the pneumococcal NVT population following the introduction of PCV-13. Additionally, we evaluate whether serotype 3 pneumococci have declined and how they have changed through this time period.

**METHODS**
Sample Collection

Pneumococcal isolates were collected from nasopharyngeal swabs of children in Massachusetts between October and April of 2000-01, 2003-04, 2006-07, 2008-09, 2010-11 and 2013-14 as previously described\textsuperscript{2,7,8}. Each sampling season is referred to by the later year. Genomes from the 2001, 2004, and 2007 sampling periods were previously published and read data for these were obtained from ENA\textsuperscript{16}. Isolates from 2009 through 2014 were sequenced from Nextera XT genomic libraries analyzed on a MiSeq.

Genomic Processing

Draft assemblies were constructed using SPAdes and annotated using Prokka\textsuperscript{18,19}. Assemblies not between 1.9 and 2.3 Mb were excluded from further analysis, as were those that produced fewer than 1900 annotations. Roary was then used to identify core and accessory genes and to generate a core gene alignment\textsuperscript{20}.

Typing

Serotype was identified using the Quellung reaction as previously described and reported for all but the 2014 sample\textsuperscript{16,17,21}. Serotypes were checked using SRST2 and a database constructed from 91 published sequences of the pneumococcal capsule biosynthetic locus\textsuperscript{22–24}.

Phylogenetic Analysis

The core genome alignment generated by Roary was used to construct a phylogeny using FastTree\textsuperscript{25}. In order to identify clusters of related sequences (SCs), three iterations of hierBAPS were run on the core genome alignment, setting the maximum cluster depth to 1 and maximum number of clusters to 30, 40, and 50\textsuperscript{26}.
Sequence Cluster Diversity

In order to determine the potential effect of PCV-13 sequence cluster diversity in this population, we calculated Simpson’s D for each sampling period. This value, which represents the probability that two randomly drawn isolates from a given sampling period belong to different SCs, was calculated as

\[
D = \frac{N}{N-1} \left(1 - \sum_{i=1}^{m} x_i^2\right),
\]

where \(x_i = \frac{n_i}{N}\), the fraction of isolates in that year belonging to sequence cluster \(i\) and \(\frac{N}{N-1}\) is a correction for finite sample size\(^{27}\).

Following an earlier analysis of serotype diversity in this population, Welch’s t test was used to compare the 2007 and 2011 populations and the 2011 and 2014 populations in order to test whether SC diversity changed following the introduction of PCV-13\(^{15}\). The polyphyletic SC was excluded from these calculations.

An increase in diversity would be expected if common lineages become rarer and rare lineages become more common. To estimate the expected change in diversity due to a smooth transition between the 2007 and 2014 population, a series of composite diversities were calculated in which the proportion of the pseudopopulation belonging to each SC was a weighted combination the 2007 and 2014 value for that SC, with the weights for the two years summing to 1. The sample size correction factor, \(\frac{N}{N-1}\), was similarly weighted.

In order to determine which SCs became more or less common following the introduction of PCV13, we conducted a Fisher’s exact test for each SC comparing its frequency between the 2007 and 2014 sample. The proportion of the population belonging to each SC and their rank order in the population were determined. As diversity increases, the shape of this distribution would be expected to flatten, with the most common lineages decreasing and the least common lineages increasing. In order to compare this distribution from the 2007 and 2014 sampling
periods with that from 2011, the frequency of each SC was plotted against its rank and overlaid with the distribution from 2011.

**NVT Composition**

To determine the clonal composition of the pre- and post-PCV-13 NVT population, the proportion of the NVT population belonging to each of the 20 SCs identified by hierBAPS was calculated for 2007, 2011, and 2014. For the purpose of these analyses, serotype 6C was considered a PCV-13 type due to its cross-reactivity with serotype 6A\textsuperscript{28,29}. Fisher’s exact test was used to determine whether these proportions varied between each pairwise combination of these three sampling periods.

We then sought to determine if the gene content of the NVT population varied between sampling periods before and after the introduction of PCV-13. Logistic models were used to evaluate the extent to which individual genes became more or less common between 2007 and 2014, as well as between 2007 and 2011. Genes were excluded if they were universally present or absent in either sampling period or present or absent in fewer than 5 total isolates between the three sampling periods. For the set of genes included in both models, we calculated a linear fit comparing the regression coefficients corresponding to the time periods from 2007 to 2011 and 2007 to 2014.

In order to determine whether changes in the gene content of the NVT population from 2007 to 2011 continued, stabilized, or reversed from 2011 to 2014, we compared the observed data to hypothetical scenarios in which the 2014 population was purely reflective of the population from either the earlier sampling periods. To do this, we drew with replacement a sample of the same size as the 2014 population from either the 2007 or 2011 population. Twenty
resampled populations were generated from each of 2007 and 2011, then used in place of the true 2014 population in the previous regression analyses. This process was repeated for an additional twenty resampled populations drawn from the true 2014 population in order to gauge its variability. This enabled us to evaluate the gene content of the 2014 population in relation to what would be expected if there was no overall change either from 2007 or from 2011.

**Evaluation of Serotype 3**

Previous studies have noted that PCV13 may not be as effective against serotype 3 as it is against the other serotypes included\(^2,11,13,14\). We compared the proportion of the pneumococcal population composed of serotype 3 between 2007, 2011, and 2014 in relation to the other three PCV-13 serotypes present in our sample, 19A, 7F, and 6C. We then used RaxML to construct a phylogenetic tree based on the core genome of serotype 3 isolates to determine if the pre- and post-PCV-13 populations were genetically distinct\(^30\).

**RESULTS**

**Sample**

A total of 1352 isolates were included in the final analysis. The core genome consists of 1000 genes found in at least 99% of isolates, producing an alignment 885 Kb in length. A total of 10,941 genes were identified. Setting the maximum number of hierBAPS clusters to 30 and 40 produced identical results, with 21 clusters identified. With the maximum number of clusters set to 50, an additional cluster was identified and another cluster was expanded. This resulted in 22 SCs, 21 of which were monophyletic and ranged in size from 14 to 177 isolates. The other, SC1, contained 150 isolates belonging to multiple small clades or individual leaves throughout the tree.
and should be interpreted as containing all lineages that could not be grouped, other than on the basis of their lack of similarity to any other cluster [Fig 2.1].
Figure 2.1: (a) Core genome phylogeny with SCs denoted by color. (b) Proportion of population in each sampling period composed of each SC, with shading indicating serotype. Solid colors are PCV-7 type, solid colors with black hatching are PCV-13, and white with colored hatching are not covered by either. Serotype 6A is dotted as it is cross-reactive with 6B, a PCV-7 type, but is itself included in PCV-13.
Diversity

Sequence cluster diversity was calculated for each year using Simpson’s D, excluding the polyphyletic cluster SC1. Diversity was significantly higher in 2011, the first sampling period following the introduction of PCV13, than it was in either 2007 or 2014, the adjacent periods for which data were available (2007 p=0.018, 2014 p=0.00098) [Fig 2.2a]. The weighted diversity estimate displayed the expected increase over either the 2007 or 2014 values, but was never as high as the diversity calculated for 2011 [Fig 2.2b].

After Bonferroni correction, only 3 SCs, SC3, SC9, and SC20, changed significantly in their share of the pneumococcal population between 2007 and 2014 (Fisher’s exact test p=0.0021, 0.0022, and 4.5x10^{-6}, respectively). SC3 became more common, increasing from 5.8% of the 2007 sample to 13.4% of the 2014 sample, with serogroups 23 and 15 coming to predominate over serogroup 6. Both SCs 9 and 20 are primarily composed of PCV-13 serotypes (7F and 6C, respectively) and were complete absent in the 2014 sample.

The overall shape of the frequency distribution was slightly flatter in 2011 as compared to 2007 and 2014, as would be expected from the higher diversity in that sampling period. Relatively rare SCs in particular were more common in the 2011 sample than the adjacent periods [Fig 2.2c,d].
Figure 2 (continued)
(Continued)

Figure 2.2: (a) Simpson’s diversity of SCs, excluding the polyphyletic cluster, for each sampling period. (b) Diversity calculated for hypothetical composites of 2007 and 2014 populations, with 2011 diversity shown as dashed line. (c-d) Proportion of population in (c) 2007 and (d) 2014 composed of each SC, ordered by frequency. Gray line is the corresponding distribution from 2011, with dotted lines representing 95% of values from 10000 random samples drawn from the 2011 population.
NVT Composition

Non-PCV-13 types increased from 66.5% of the pneumococcal population in 2007 sampling period to 92.3% in the 2014 sampling period. Fifteen SCs had at least 1 NVT isolate. There was no significant difference between the SC distribution amongst NVTs in 2007 and 2014 (Fisher’s exact test p=0.24). There was, however, a significant difference between 2007 and 2011 (p=0.0018) and between 2011 and 2014 (p=0.0078), indicating that the population was disrupted in 2011 but returned to its pre-vaccination state by 2014. Correspondingly, the proportion of the NVT population comprised of several of the most common SCs showed a distinct increase from 2007 to 2011 and decrease from 2011 to 2014, or vice versa [Fig 2.3].

Figure 2.3: Proportion of the NVT population comprised of each SC. Two additional SCs, SC11 and SC20, had a single NVT isolate and were excluded from this plot.
This is partially reflected by the trend in gene content over time. The linear fit comparing the 2007-2011 and 2007-2014 regression coefficients for each gene had a slope of 0.62, indicating less overall change between 2007 and 2014 than between 2007 and 2011. This slope fell between those from hypothetical 2014 populations drawn from either 2007 or 2011, which clustered around a slopes of 0 and 1, respectively [Fig. 2.4]. This indicates that while the direction in which genes changed in frequency from 2007 to 2011 was generally preserved through 2014, the trend was partially counteracted between 2011 and 2014 with genes returning closer to their 2007 levels.

**Figure 2.4:** Regression coefficients comparing gene content of the NVT population from 2007 to 2011 and 2014. Black circles correspond to the coefficients with individual genes, with a linear fit to the data shown in black. Fits in which a hypothetical 2014 population was drawn from either the 2007, 2011, or 2014 population are shown in red, blue, and gray, respectively.
Persistence of Serotype 3

In order to evaluate whether the new serotypes included in PCV13 decreased following its introduction, we conducted a Fisher’s exact test comparing the 2007 and 2014 carriage share of serotypes 19A, 6C, 7F, and 3. While serotypes 19A, 6C and 7F all showed significant reductions between the two time periods (p<0.0001, p=0.00014, p=0.0011, respectively), serotype 3 had no such change (p=0.46) [Fig 2.5].

Figure 2.5: Proportion of the population in each sampling period comprised of the serotypes included in PCV-13 but not PCV-7
To test if the persistence of serotype 3 may be related to some genetic factor, Roary was run using only the serotype 3 sequences and a RaxML was used to construct a phylogeny based on the core genome. While all isolates clustered into the same SC and share the same MLST profile, there is a distinct bifurcation in the tree [Fig 2.6]. Of the 28 serotype 3 isolates, 16 fall into one subclade and 12 into the other. In the larger subclade, 4 isolates (25%) are from 2011 or 2014, after the introduction of PCV-13. The other subclade contains 11 (92%) post-PCV-13 isolates ($\chi^2 p=0.0018$).

**Figure 2.6:** Serotype 3 phylogeny, with sampling period shown by color. Isolates collected before the introduction of PCV13, shown in blue, are found primarily on one branch of the tree, while post-introduction isolates, indicated by red, are primarily on the other.

**DISCUSSION**

Here we have analyzed a sample of carriage pneumococci collected in Massachusetts between the winters of 2000-01 and 2013-14, focusing primarily on changes occurring following the introduction of PCV-13. Using genomic data, we find that the NVT population in the most
recent sampling period more closely reflects that of our last full pre-PCV13 sample than our first post PCV-13 sample. We also find that serotype 3 has been more persistent than other serotypes added for PCV-13, but a different subclade of this lineage has come to predominate.

Given the value of being able to predict the composition of the pneumococcal population following PCV use, the pattern observed amongst the NVTs is quite interesting. Our 2014 sample appears to be broadly a reflection of the 2007 sample, but 2011 is unlike either. As such, it is possible that the pre-vaccine NVT population may be a good predictor of the post-vaccine population, but that the disruption caused by vaccine introduction can temporarily interrupt this pattern. Some of this could be due to variation in the age of children who have been vaccinated, which should increase over time as vaccinated children age. The observed increase in sequence cluster diversity in the immediate post-vaccine period, with the most common lineages making up a smaller proportion of the total population, may provide an enhanced opportunity for rarer lineages to become more increase. Lineages such as SCs 3, 14, and 19 (serotypes 23A/15BC, 21 and 33F, respectively), may have a similar trajectory to that of serotype 19A ST320 after PCV-75,9,10,31. Further observation will help determine if these or other lineages become more substantial contributors to both carriage and invasive disease.

Previous studies have indicated that PCV-13 may not be as effective against serotype 3 as it is against other included serotypes2,11,13,14. The shift we observed in the serotype 3 population following the introduction of PCV-13 may reflect a similar phenomenon to that leading to the recognition of serotype 6C as distinct from 6A following the introduction of PCV-732,33. The dominant lineage pre-PCV-13 was also more homogenous than the post-vaccination population, so it is possible that the immunity generated by to serotype 3 by PCV-13 is narrowly tailored to
that subset of the population. Given its propensity for causing disease, the persistence of serotype 3 despite its inclusion in PCV-13 warrants further investigation.

The response of the pneumococcal population to serotype-targeting conjugate vaccines may also provide insights for other pathogens for which vaccines have been targeted at or differentially affect a subset of their population. The efficacy of the RTS,S vaccine appears to be partially dependent on how well the circumsporozoite protein of a given malaria parasite matches that in the vaccine. There has also been interest in understanding how the strain dynamics and epidemiology of meningococcal disease will be affected by the rollout of vaccinations against a variety of serogroups. While each of these disease systems is different, there is some potential for findings in one to inform hypotheses for how others will behave.

Pneumococcal epidemiology has changed substantially as a result of conjugate vaccination. While PCVs have been highly effective in reducing the incidence of pneumococcal disease, continued vigilance is necessary to monitor for and respond to the emergence of potentially dangerous lineages not protected against by current vaccine formulations.


Chapter 3: Comparison of Carriage and Invasive Serotype 33F Pneumococci Before and After PCV-13 Introduction

INTRODUCTION

*Streptococcus pneumoniae* (the pneumococcus) is a common bacterial pathogen responsible for a substantial burden of disease. In most cases, however, pneumococcal colonization results in nothing more than asymptomatic carriage. Pneumococci are commonly classified into one of more than 90 different serotypes, defined by the reactivity and structure of their polysaccharide capsule. The burden of disease due to the different serotypes is a function of both their overall prevalence in the population and their propensity to cause disease once acquired, which are highly variable. As such, some serotypes cause a high burden of disease by virtue of being very common, while rarer serotypes may cause a similar amount of disease by sickening a greater proportion of their carriers.

Pneumococcal conjugate vaccines have been highly successful in reducing invasive pneumococcal disease (IPD) despite constant levels of carriage by targeting the serotypes responsible for the greatest amount of disease. Small increases in disease due to non-vaccine types (NVTs) were observed, however, following the introduction of the heptavalent PCV (PCV-7). The emergence of serotype 19A as a significant cause of pneumococcal disease associated with non-susceptibility to multiple classes of antibiotics was of particular concern. Similar effects could occur in the wake of PCV-13, which was introduced in the United States in 2010.

Amongst the serotypes not included in PCV-13, 33F has one of the highest degrees of invasiveness – defined as the incidence of invasive disease from the serotype relative to its carriage prevalence. In the United States, it was noted as one of the primary IPD replacement
serotypes in following the introduction of PCV-7, and was the most common IPD serotype in 2010 and 2011, following the introduction of PCV-13\textsuperscript{20–23}. Similar increases in IPD due to serotype 33F have been observed in Canada and the United Kingdom\textsuperscript{24,25}. In England between 2002 and 2011, 33F caused the greatest loss in QALYs in children under age 5 amongst non-vaccine types\textsuperscript{26}.

Serotype 33F includes pneumococci from several distinct lineages as defined by pulse-field gel electrophoresis (PFGE) and multi-locus sequence typing (MLST). The relative abundance of these different types is dynamic, varying temporally and geographically\textsuperscript{20,21,24,25,27–30}. In addition to differences on the basis of these typing schemes, a variety of resistance phenotypes have been observed amongst 33F pneumococci, with macrolide resistance in particular being identified in multiple studies\textsuperscript{20,24,29,31}. Serotype 33F pneumococci have not, however, been thoroughly studied using the greater resolution available through whole genome sequencing. Fluctuations at the serotype level could be due to a single lineage or could comprise multiple lineages with either concordant or discordant patterns. While MLST data can partially overcome this, it can still miss finer scale changes in the makeup of a particular sequence type.

Here we examine a collection of 33F pneumococci from Massachusetts, comprised of both carriage and invasive disease isolates sampled from shortly after the introduction of PCV7 through several years after the introduction of PCV-13. Using short read sequencing data, we describe the lineages comprising this population over time, evaluate its invasive capacity, and compare the genetic composition of carriage and invasive isolates.

**METHODS**

**Sampling**
The sampling of carriage isolates has been previously described. In brief, isolates were collected from nasopharyngeal swabs of children under the age of 7 visiting a participating healthcare provider in Massachusetts during the winters ending in 2001, 2004, 2007, 2011, and 2014. Those identified as belonging to serotype 33F were included in this study. The total number of isolates for each sampling period was taken as those included in Chapter 1 of this dissertation. Invasive serotype 33F isolates from children in this age range were taken from a previously described statewide enhanced surveillance system for pediatric pneumococcal disease. Only invasive isolates collected between 2003 and 2014 were included.

**Genetic Characterization**

Serotype was verified and multi-locus sequence type (MLST) data were obtained prior to assembly using SRST2. SPAdes was used to construct de novo draft assemblies, which were then annotated using Prokka. Roary was then used to identify homologous genes between isolates and identify core and accessory genes. RaxML was used to construct a phylogeny based on an alignment of core genes. Given previous reports of non-susceptibility to macrolides and other antibiotics in 33F pneumococci, we used BLAST to search the pan genome for \( \text{ermB} \) and the \( \text{mel-mef} \) efflux pump, which confer high and low level macrolide resistance, respectively. We also searched for \( \text{tetM} \) which confers tetracycline resistance and frequently carried with \( \text{ermB} \).

**Temporal trends**

To evaluate whether serotype 33F has become a more common cause of pneumococcal colonization following the introduction of PCV-13, we calculated the proportion of isolates collected that were identified as 33F in each carriage sampling period. We also sought to
determine whether the clonal composition of 33F pneumococci in our population varied over time. To do this, we grouped isolates into clonal complexes (CCs), groups of MLST profiles with identical alleles at 6 of the 7 MLST loci. We then used Fisher’s exact test to determine whether the distribution of CCs varied between the pre- and post-PCV-13 periods. For invasive isolates, those collected between 2003 and 2009 were considered pre-vaccine and those from 2010-2014 post-vaccine. For carriage isolates, the 2001, 2004, and 2007 sampling periods were pre-vaccine and 2011 and 2014 were post-vaccine.

**Invasive Capacity**

In order to determine whether serotype 33F pneumococci retained their high propensity for causing invasive disease, we calculated invasive capacity (IC) for the pre- and post-PCV-13 samples. Following a previous study of multiple serotypes in this population, this was defined as

\[ IC = \frac{I_x}{q_x} \]

where \( I_x \) is the incidence of invasive disease and \( q_x \) is the proportion of carriage swabs positive for serotype 33F, providing an estimate of carriage prevalence. \( I_x \) is calculated as the number of cases of 33F IPD divided by 7, the number of years over which isolates were collected, and divided by the estimated number of children aged <7 years in Massachusetts during that time frame, providing an estimate of serotype-specific annual incidence per capita. For the pre-vaccine period, the population was taken as 541,887.5, the mean of the 2000 and 2010 U.S. census counts. For the post-vaccine period, the 2010 census count, 519,528, was used and the sampling period was 5 years. The natural logarithm of the IC is normally distributed with variance \( \frac{1}{N_x} + \frac{1-q_x}{q_x m} \), where \( N_x \) is the population size estimate and \( m \) is the number of carriage swabs collected for the relevant time period. These calculations were done for both the total 33F population and for individual clonal complexes.
**Genetic Distinction between Carriage and Invasive Isolates**

We sought to determine whether there was any genetic distinction between carriage and invasive isolates. Pairwise genetic distances were extracted from the core genome alignment using the R package SeqinR\textsuperscript{43,44}. Distance was defined as $\sqrt{1 - i}$, where $i$ is the identity between two sequences. As an omnibus test of whether carriage and invasive isolates tended to cluster together, we calculated an estimate of $F_{ST}$ using the formula $\frac{\pi_B - \pi_W}{\pi_B}$, where $\pi_B$ is the mean core genome identity between pairs of isolates in which both are either carriage or invasive and $\pi_W$ is the same measure for pairs of isolates in which one is from carriage and the other is from invasive disease\textsuperscript{45}. Significance was assessed by comparing the calculated value to 10,000 permutations. To explore this relationship in finer detail, we plotted the proportion of pairs within a series of threshold distances in terms of core genome divergence and gene content for which both isolates came from either carriage or invasive disease compared to a null expectation based on 100 permutations.

**RESULTS**

**Sample Collection**

The final sample included 27 carriage and 41 invasive isolates. One invasive isolate was excluded due to poor sequence quality. Another invasive sample was excluded when it was genotyped as a serotype 10A strain and subsequently found to have been a mixed culture.

**Genetic Characterization**

A total of 2649 genes were identified, 1597 of which were core genes present in all isolates. These core genes were concatenated into a 1.47 Mb alignment.
Each isolate was assigned to a sequence type (ST) and clonal complex (CC) based on sequence reads and a phylogeny was constructed based on a core genome alignment. A total of four clonal complexes were identified, one of which is separated from the others by a long branch [Figure 3.1]. Two of these clonal complex, CC717 and CC1012, were not identified in our carriage sample until after the introduction of PCV-13, although they contributed to the invasive disease sample in earlier years [Table 3.1].

Table 3.1: Carriage and Invasive Isolates by Clonal Complex

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CC100</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>CC2705</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CC717</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>CC1012</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

| Total Isolates | 133 | 203 | 278 | 365 | 313 |
| Total Swabs⁷   | 678 | 987 | 1540| 1164| 1157|

<table>
<thead>
<tr>
<th>Invasive</th>
<th>2003-09</th>
<th>2010-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC100</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>CC2705</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>CC717</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CC1012</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>17</td>
</tr>
</tbody>
</table>

Within this population the two macrolide resistance elements, *ermB* and the *mel-mef* efflux pump were both present but segregated from one another. In total, 45 isolates showed evidence of one of these genes The 9 isolates in which *ermB* was identified were those in CC717.
As expected, these isolates also carried *tetM*. Both genes corresponding to the *mel-mef* efflux pump were found in 36 isolates, including all of those in CC2705 and some in CC100 [Fig 3.1].
Figure 3.1: Core genome tree of 33F isolates. CCs 717 (orange), 2705 (blue), and 1012 (green) are highlighted. All other leaves belong to CC100. The first bar indicates whether the leaf is a carriage (blue) or invasive (red) isolate. The second bar indicates the presence of \textit{ermB} and \textit{tetM} (orange) or the \textit{mel-mef} efflux pump (blue).
Increasing Carriage

Serotype 33F pneumococci markedly increased amongst carriage isolates between the 2001 and 2014 sampling periods. Between 2001 and 2007, 33F pneumococci never accounted for more than 1% of carriage isolates collected in a sampling period, but increased to 2.2% by 2011 before doubling to 4.5% by 2014 [Table 3.1]. This indicates that, following the introduction of PCV-13, 33F has become an increasingly common carriage serotype, allowing for a greater chance to cause disease.

Declining Invasive Capacity

As serotype 33F has previously been reported to be amongst the most invasive serotypes not covered by PCV-13 and has become increasingly common in carriage, we estimated its invasive capacity pre- and post- vaccine introduction. The invasive capacity calculated over our entire sample declined significantly, from $4.1 \times 10^{-3}$ (95% CI: $1.5 \times 10^{-3} - 1.1 \times 10^{-2}$) to $6.9 \times 10^{-4}$ (95% CI: $3.7 \times 10^{-4} - 1.3 \times 10^{-3}$) between the pre- and post-PCV-13 periods ($z=3.01$, $p=0.0026$). This decline is reflected, though with only marginal statistical significance, when focusing specifically on CC100, which declined from $3.4 \times 10^{-3}$ (95% CI: $9.5 \times 10^{-4} - 1.2 \times 10^{-2}$) to $7.3 \times 10^{-4}$ (95% CI: $3.0 \times 10^{-4} - 1.8 \times 10^{-3}$) ($z=1.95$, $p=0.051$). The invasive capacity of CC2705 did not change significantly, measuring $3.0 \times 10^{-3}$ (95% CI: $6.1 \times 10^{-4} - 1.4 \times 10^{-2}$) before the introduction of PCV-13 and $1.1 \times 10^{-3}$ (95% CI: $3.0 \times 10^{-4} - 4.3 \times 10^{-3}$) after ($z=0.93$, $p=0.35$). The other two CCs, CC717 and CC1012, were not present in our carriage collection until the 2011 sampling period, so no pre-PCV-13 IC was calculated. The estimated post-PCV-13 IC for these CCs were slightly
lower than those for the larger CCs, although with wide confidence intervals (CC717: $3.6 \times 10^{-4}$, $6.9 \times 10^{-5} - 1.8 \times 10^{-3}$; CC1012: $4.4 \times 10^{-4}$, $4.1 \times 10^{-5} - 4.9 \times 10^{-3}$) [Fig. 3.2].

**Figure 3.2:** Invasive capacity for the overall 33F population and each clonal complex before (left) and after (right) the introduction of PCV-13, with error bars indicating 95% confidence bounds. For CCs 717 and 1012, only a post-introduction IC was calculated.

**Genetic Similarity between Carriage and Invasive Isolates**

In order to determine whether there was a gross genetic distinction between carriage and invasive 33F pneumococci, we compared both the core genome distance and gene content divergence within and between the two groups. By both metrics, there is no evidence of large scale genetic differentiation between carriage and invasive isolates [Fig. 3.3a, b]. For the core genome, the mean identity was 99.80% both within and between groups, corresponding to an $F_{ST}$
of -0.031 (p=0.43), indicating no discernable difference between the carriage and invasive samples. Similarly, the Jaccard distance of gene content was 0.083 within groups and 0.084 between groups. When comparing at specific distance thresholds, however, there is some evidence that isolates from the same source cluster together at genetic distances less than 0.006 [Fig. 3.3c, d]. Of the 84 isolate pairs from the same source within this threshold, 55 (65.5%) were invasive, as compared to 820 of the 1171 total within-group pairwise comparisons (70.0%), indicating that the balance of carriage and invasive within-group comparisons amongst closely related pairs was similar to that of all within-group pairs.
Figure 3.3: Overlapping histograms of the genetic relatedness for pairs of isolates in which one is carriage and the other invasive (between) and where both are from the same source (within), expressed as core genome identity (a) and Jaccard distance of gene content (b). Proportion of isolates from the same source within thresholds of genetic relatedness. Red circles are the result of 100 permutations, with blue lines representing the 2.5th and 97.5th percentiles of permutation results (c and d).
DISCUSSION

In this study we have examined the changing population structure and invasive capacity of a single pneumococcal serotype, 33F in the years surrounding the introduction of PCV-13. This serotype has previously been identified as having a particularly high degree of invasiveness and being responsible for a substantial burden of disease relative to other serotypes not included in PCV-13 in multiple locations, and is widely found in carriage at low frequencies.\textsuperscript{6,19,26} Hence there is justifiable concern that any increase in 33F carriage following vaccination may lead to more 33F IPD. Here, we find evidence that the invasive capacity of serotype 33F may have decreased in the years following the introduction of PCV-13.

There are several possible explanations for this observed decline in invasive capacity. One is a shift in the population carrying 33F. As PCV-13 reduces the carriage prevalence of VT pneumococci, NVTs will take their place\textsuperscript{7}. We have observed an increase in the carriage prevalence of serotype 33F in the post-PCV-13 era, suggesting that it is taking advantage of the carriage space previously occupied by VTs. This could result in an apparent decrease in invasive capacity if the new host range is less susceptible to progressing from carriage to disease, for example because of a changing age distribution within the <7 year old age group, as invasive disease incidence is known to vary dramatically within this age group, peaking around 1 year\textsuperscript{6,46}. This would likely manifest as a lower IC across 33F lineages, which our data suggest but do not provide the power to thoroughly investigate. Two other potential explanations are supported by our finding that very closely related pairs of isolates have an increased probability of both being either carriage or invasive samples. The first of these is that the observed decline in IC is a sampling artifact arising from the difference in how carriage and invasive isolates were collected.
While invasive isolates were collected continuously, carriage isolates were only collected during discrete time periods. Other pneumococcal lineages have been shown to have a cyclical temporal trends\textsuperscript{47}, so it is possible that our carriage sampling may have differentially sampled during high and low periods for 33F between the pre- and post-PCV13 periods. Alternatively, it is possible that there has been a true shift in the population composition of serotype 33F, with less invasive sub-lineages taking precedence. This is supported by the dense clustering of carriage isolates from 2011 and 2014 with one another rather than intermingled with invasive isolates. Further study of serotype 33F as well as other lineages could help determine the true underlying cause or causes of our observed decrease in invasive capacity.

While their role in invasive capacity is uncertain, two new clonal complexes have become part of our 33F population in recent years. Records of ST1012 and ST673, the two STs comprising CC1012 in our sample, can be found linked to a variety of serotypes in the pneumococcal MLST database (pubmlst.org), with a 2003 isolate from Scotland labeled as serotype 33. The earliest isolate explicitly labeled 33F was collected in Germany in 2007. This clonal complex first appeared in our invasive sample in 2003 but did not appear in carriage until 2011. CC717 entered our population more recently, with the first invasive isolate collected in 2007 and not appearing in carriage until 2011 but representing 29% of the 2014 carriage sample. The expansion of this CC is of particular concern as it harbors $ermB$, which confers high level macrolide resistance. Serotype 33F pneumococci of this lineage have previously been linked to macrolide resistance in Europe since the late 1990s\textsuperscript{27,29}. The presence of the $mel$-$mef$ efflux pump in CCs 100 and 2705, however, may mean that an increase in the prevalence of $ermB$ amongst 33F pneumococci would have little practical impact on treatment options.
As the composition of the pneumococcal population shifts with the expansion of PCV use, both in valence and geographic coverage, it is important to monitor which lineages could become significant contributors to the burden of pneumococcal disease. Our observed decrease in the relative invasiveness of 33F is encouraging, given that it has been associated with a particularly high propensity for causing disease. Further study of this and other pneumococcal subtypes will help maximize the benefit of the current strategy of vaccinating against particular serotypes until more broadly target interventions become feasible.

REFERENCES


CONCLUSION

The early years of the 21st century have been a time of significant change in pneumococcal epidemiology. Pneumococcal conjugate vaccines have been a powerful tool for reducing the burden of pneumococcal disease, though their selective nature means that NVT pneumococci have expanded to fill the void left by those targeted by vaccination1–6. This strategy has still been highly successful, as pneumococci vary in their propensity to cause disease once acquired7,8. However, the emergence of new or previously rare lineages with high invasive capacity or other clinically significant traits, such as antibiotic resistance, remains a concern.

Chapter 1 examined some factors that may contribute to the emergence of such a lineage. While focused on antibiotic resistance, it is applicable to other traits as well. The results if this model suggest that sufficiently strong selective pressure in favor of a trait can overcome vaccine pressure against it, and that the potential for this emergence is greatest in setting of high carriage burden and multiple colonization. These are typically found in resource limited settings, where vaccine delivery may already be hindered and surveillance capacity may be lacking9,10. The role of multiple carriage is compounded by the potential it creates for recombination to produce novel variants.

Chapter 2 helps us understand the relationship between the pre- and post-vaccination pneumococcal landscape. By and large, the distribution of NVTs prior to PCV-13 introduction mirrors the present population. However, the 2011 population was quite different in terms of the balance of lineages. Similarly, while changes in gene content from 2007 to 2011 were strongly predictive of the total change from 2007 to 2014, they were partially counteracted by trends from 2011 to 2014. Studies in other populations would help elucidate whether the effects observed here are a general feature that can be expected after vaccine introduction or simply a chance
fluctuation in our sample. It is possible that post vaccination shuffling of the population may provide opportunity for rarer lineages to gain ground. Continued observation and simulation work could help clarify the importance of this period for emergence.

Chapter 3 considers the changes within a single serotype, 33F. This serotype was previously recognized as one of the most invasive NVTs and, as such, warrants close monitoring as the pneumococcal population moves towards a new equilibrium\(^7,8\). The invasive capacity of this serotype appears to have declined following the introduction of PCV-13. This may reflect several factors, such as a change in host range induced by vaccination or shifts in the genetic composition of the lineage. More broadly, it shows that the characteristics of a given serotype are not necessarily static and a lineage may behave differently in a vaccinated population. This suggests a need for similar studies for other NVTs, as it is unclear how well pre-vaccine measurers of invasive capacity predict those after vaccine introduction. This may have implications for future serotype specific control strategies.

Taken as a whole, this dissertation contributes to our understanding of how the pneumococcal population has and will continue to react to the introduction and expansion of PCV use. Serotype specific conjugate vaccines have been successful in reducing the serotypes they target, and the remaining pneumococcal population is generally reflective of the NVT population prior to vaccine introduction. However, certain characteristics of these lineages may be altered. Furthermore, there may be a period of disruption that obscures this pattern and provides an opportunity for the emergence of vaccine escape lineages. So long as pneumococcal vaccination remains focused solely on serotype, it will be important to consider these factors when designing and implementing updated vaccination strategies. Even if more broadly immunizing, non-serotype specific pneumococcal vaccines can be successfully developed,
similar forces may be at play for any vaccine that targets only a subset of the population of a given pathogen.


