UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Clinical and Experimental Sciences

Volume [1] of [1]

PNEUMOCOCCAL CARRIAGE AND DISEASE DURING THE IMPLEMENTATION OF PCV13: 2006 to 2016

by

Jessica Jones

Thesis for the degree of Doctor of Philosophy

March 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Molecular Microbiology

Thesis for the degree of Doctor of Philosophy

PNEUMOCOCCAL CARRIAGE AND DISEASE DURING THE IMPLEMENTATION OF PCV13: 2006 to 2016

Jessica Jones

Streptococcus pneumoniae is both a common inhabitant of the human nasopharynx and a leading cause of morbidity and mortality. Characterisation of pneumococcal populations derived from asymptomatic carriage and invasive disease during the introduction of a 13-valent pneumococcal conjugate vaccine (PCV13) allows the impact of the vaccine to be assessed and may provide information with which to inform public health policies.

Streptococcus pneumoniae isolated from the nasopharynx of healthy children were collected alongside isolates found in invasive pneumococcal disease (IPD) from the same geographical location over a ten year period. Carriage rates and IPD incidence rates were examined to assess the impact of PCV13 on both populations. Changes to the serotype distribution in both the carriage and the disease populations were compared and a molecular investigation was performed to ascertain similarities between the populations.

Pneumococcal carriage and IPD incidence caused by PCV13 serotypes decreased following PCV13 introduction however an increase in carriage and IPD incidence caused by non-vaccine serotypes was observed. Evaluation of serotype distribution revealed that three non-vaccine serotypes had significantly higher odds of featuring in IPD; 12F, 8 and 9N. Genotypic analyses showed that while 83% of IPD isolates had the same clonal type as carriage isolates, some molecular variations within prevalent carriage isolates were not seen in the IPD population.

The increase in carriage and IPD caused by non-vaccine serotypes is compromising the benefits of PCV use. Continued surveillance is required to safeguard public health.

Table of Contents

Table of Tables

Table of Figures

Academic Thesis: Declaration Of Authorship

I, Jessica Jones declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

'Pneumococcal carriage and disease during the implementation of PCV13: 2006 to 2016'

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: .....

Date: .....

Study Contributions

The paediatric pneumococcal carriage study is an on-going study based at Southampton General Hospital. Year one is dated from the winter of 2006/07 and subsequent study years cover every winter until year 10, winter 2014/15. Analyses have been carried out by three former PhD students: Dr Anna Tocheva; Dr Rebecca Gladstone and Dr Vanessa Devine. Publications derived from the carriage study data are listed in Appendix A.

My contribution to the study began in 2013/14 and the work undertaken in the production of this thesis is detailed below:

- Serotype analysis of all carriage isolates collected between 2013/14 and 2015/16
- Molecular characterisation of all carriage isolates collected between 2013/14 and 2015/16
- Additional serotyping and molecular typing of isolates collected prior to 2013/14 that had missing data at the start of my candidature.
- Collation and statistical analysis of all carriage data collected between 2006/07 and 2015/16
- Microbiological identification of isolates collected between 2013/14 and 2015/16 (alongside employed laboratory technicians).
- DNA extraction of all carriage isolates collected between 2013/14 and 2015/16
- Retrieval and DNA extraction from all IPD isolates collected between 2006/07 and 2015/16.
- Collation and analysis of metadata relating to the IPD isolates collected between 2006/07 and 2015/16
- Retrieval and storage of whole genome data for all carriage and IPD isolates collected between 2006/07 and 2015/16.
- All bioinformatic analyses contained within this thesis

The carriage study relied upon external contributors for some aspects of the study. Any work carried out by external contributors is clearly indicated within this thesis.

Table 1 Contributions to analysis of carriage study isolates

Study Year	Winter	Analysis	Performed by
1	2006/07	Pneumococcal typing/MLST	Dr Anna Tocheva
2	2007/08	Pneumococcal typing/MLST	Dr Anna Tocheva
3	2008/09	Pneumococcal typing/MLST	Dr Anna Tocheva
4	2009/10	Pneumococcal typing/MLST	Dr Rebecca Gladstone
5	2010/11	Pneumococcal typing and analysis of whole genome sequences (WGS) of years 1-5	Dr Rebecca Gladstone
6	2011/12	Pneumococcal typing and analysis of WGS of year 6 isolates	Dr Vanessa Devine
7	2012/13	Pneumococcal typing and analysis of WGS of year 7 isolates	Dr Vanessa Devine
8	2013/14	Pneumococcal typing and analysis of WGS of year 8 isolates	Mrs Jessica Jones
9	2014/15	Pneumococcal typing and analysis of WGS of year 9 isolates	Mrs Jessica Jones
10	2015/16	Pneumococcal typing and analysis of WGS of year 10 isolates	Mrs Jessica Jones

Acknowledgements

I would first like to thank my supervisory team. I have been fortunate in having guidance from five experts that I respect and admire. Saul Faust and Jo Jefferies give my research a real world perspective that makes me step back and think of the wider implications of the work that I do. Rebecca Gladstone has been invaluable in her knowledge both of the subject and of this study. Her steady advice and her ability to understand what I need have assisted me at the points when I've needed help most. David Cleary's calm day to day guidance from what analytical tools to use to how to balance work:life has been invaluable. He has often been the person that I have vented to and his unique humour and steady influence have bolstered me throughout this process (I still won't use the word 'pathobiont' though David). Special thanks to my main supervisor, Stuart Clarke. For the last four years I have been fortunate to receive Stuart's advice and guidance. He has understood when and how to help and provided me with a myriad of opportunities to progress in academia. Whether he is just down the corridor or on the other side of the globe I have always had his support and I could not have undertaken this research without him.

Huge thanks also go to my colleagues. Not only have they supported my research but we have developed important friendships that will last beyond the work that we do. Rebecca, Abi, Vanessa and Elita balance professionalism with great humour. Karen has been one of the most reliable people in my life. Her wit, intelligence and generosity have helped me through a number of episodes and I can't thank her enough. Also thanks to the technicians and microbiologists that have supported this study especially Andy Tuck who gave me my first break and who I am blessed to have learned from.

A special thank you to Denise Morris. Her efforts and diligence have shaped this study. From recruitment to lab work to tracking down nurses she has managed the day to day activities and her hard work has afforded me the luxury of concentrating on my research. I can't thank her enough for all that she has done.

I must thank my friends Lucy Grave and Laura Cornwell for the many times they have stepped in at a moment's notice to grab my kids from school when I've got stuck in the lab or overslept after an all-nighter. I couldn't have managed without them both and my family is fortunate to have such amazing women in our lives.

Finally, thanks to my family: my mother, who gave me the self-belief to tackle such a thing as this in the first place, my father for his encouragement and kind words and, my sister who is always there to offer support and make me laugh. From organising Christmas to taking care of the kids they have ensured that life has carried on as normal and provided sanctuary whenever it was needed.

Above all else I offer heartfelt thanks to my husband Kevin and my sons, Monty and Reuben. We have truly made this journey together and I am so proud of the four of us. Kev, what can I say? Your love and generosity have amazed me. Thank you so much for all the help you've given me. For the last few months you have carried our family and allowed me to concentrate on my research. And, to my beautiful boys, I know the last couple of months have been frustrating at times and you've shown amazing understanding when I've been preoccupied. Yes, Mumma has finished her work now. I hope one day it makes you proud.

Kevin, Monty and Reuben, I dedicate this thesis to you.

Definitions and Abbreviations

aa .....Amino acid

ATCC .....American type culture collection

ATP .....Adenosine triphosphate

BHI .....Brain heart infusion

BLAST .....Basic local alignment search tool

bp ..... Base pair

CBP ..... Choline-binding protein

CC ..... Clonal complex

CDS ..... Coding sequences

CSF ..... Cerebrospinal fluid

DNA ..... Deoxyribonucleic acid

dNTP ..... Deoxynucleoside triphosphate

IPD ..... Invasive pneumococcal disease

MLEE .....Multilocus enzyme electrophoresis

MLST .....Multilocus sequence typing

PCR .....Polymerase chain reaction

PCV ..... Polysaccharide conjugate vaccine

PFGE ..... Pulsed field gel electrophoresis

PMEN .....Pneumococcal molecular epidemiology network

SNP ..... Single nucleotide polymorphism

ST.....Sequence type

Chapter 1 Introduction

1.1 What is a pneumococcus?

A review of literature pertaining to the bacterium *Streptococcus pneumoniae* (the pneumococcus) inevitably reveals landmark moments in the biological sciences. The eminent, Robert Austrian who revealed and collated much of what we know about the pneumococcus, claimed that research surrounding this tiny organism was responsible for the emergence of such branches of science as modern immunology and molecular genetics (Austrian, 1981).

For over a century, researchers have discovered and confirmed many of the complex qualities that characterise this interesting bacterium. Its paradoxical existence as both an asymptomatic commensal and a leading causative agent in human disease has driven many scientific discoveries and kept bacteriologists keenly focused (Weiser, 2010).

As one of the most intensely studied organisms in scientific history, there now exists a rich array of information that researchers can draw from. Despite this wealth of knowledge, many questions remain unanswered and present day scientists continue to make fresh discoveries or dispel old assumptions.

Though written in 1938, the following paragraph remains relevant in 2017:

> "The study of the members of this small group of microorganisms in a subordinate branch of biology is bringing light into some of the obscure realms of the related sciences. The peculiarities of Pneumococcus are yielding a generous return to the investors and speculators who have cast in their resources with its lot, resulting in the accumulation of a store of solid bullion for the scientist and for mankind (White, 1938)."

1.1.1 The discovery of the pneumococcus

Characterisation of the pneumococcus is conventionally charted from 1881 when Pasteur and Sternberg, working independently of each other, documented descriptions of the organism (Pasteur, 1881, Sternberg, 1881).

Following observations of coccoid bacteria in human saliva: Sternberg in his own saliva and Pasteur in the saliva of child who had died of rabies, both men subcutaneously inoculated rabbits with infected saliva. The subsequent recovery of diplococci from the animals' blood demonstrated for the first time that an organism carried asymptomatically can have pathogenic potential.

Contemporaneously, Friedlander, Fraenkel and Weichselbaum working separately on the examination of lung disease documented the pneumococcus as an aetiological agent of human pneumonia (Weichselbaum, 1886, Fraenkel, 1884, Friedlander, 1883). Hans Christian Gram working in Friedlander's laboratory was able to visualise the pneumococcus while experimenting with differential staining techniques on sections of lung tissue taken from patients who had died of pneumonia (Gram, 1884). His eponymous Gram stain is still used in laboratories today.

Investigations into other disease states revealed that pneumococcal pathogenicity is not limited to pneumonia. In 1887, Netter showed that the pneumococcus was the causative agent in cases of meningitis (Netter, 1887) and Zaufel isolated pneumococci from the inner ear of both asymptomatic patients and those suffering with acute otitis media (Zaufal, 1887). Importantly both Netter and Zaufel documented that pneumococcal presence was found without accompanying lung infection thus dispelling a common belief of that time that pneumococcal infection always disseminated via the lungs.

By the end of the 1880's the pneumococcus had been identified, visualised and its role in disease was beginning to be revealed.

1.1.2 Taxonomy

Historically the pneumococcus has been known by a number of names. On discovery Pasteur referred to the organism as *Microbe septicemique du salive* (Pasteur, 1881) while Sternberg named it *Microcroccus pasteuri* (Sternberg, 1885) in homage to Pasteur. It was Fraenkel that first referred to it as Pneumococcus due to its association with lung disease (Fraenkel, 1886), however it was Weichselbaum's designation *Diplococcus pneumoniae* that became the most common way of referring to the organism until it was classified as belonging to the streptococcal genus in the 1970s (Wannamaker and Matsen, 1972). From 1974, the organism has been called *Streptococcus pneumoniae* (Buchanan and Gibbons, 1974). The genus is characterised as being Gram-

positive, catalase-negative facultative anaerobes which forms distinctive chains *in vitro*. The chains of the pneumococcus are usually so short that cocci are frequently seen as just pairs, hence the original 'diplococci' nomenclature.

1.1.3 Biological processes of the pneumococcus

Streptococci belong to the lactic acid bacterial group and as such their energy is derived from the catabolism of carbohydrates (Klaenhammer *et al.*, 2005). Adenine triphosphate (ATP) is produced during the fermentative breakdown of sugars which first follows the classical Embden-Meyerhof-Parnas pathway to generate pyruvate and nicotinamide adenine dinucleotide, NADH (Gaspar *et al.*, 2014). Pyruvate conversion can follow one of two paths: homolactic or mixed-acid fermentation, the selection of which is dependent upon environmental factors such as the abundance of oxygen or sugar. The end product of homolactic fermentation is mainly lactic acid while mixed-acid fermentation can generate other products such as acetate and ethanol (Ramos-Montanez *et al.*, 2010).

Although preferentially anaerobic, pneumococci are bestowed with a number of unusual attributes which allow it to tolerate aerobic conditions (Yesilkaya *et al.*, 2013). It has been shown to produce high quantities of hydrogen peroxide, H_2O_2, when exposed to O_2 and yet it does not produce catalase as other bacteria usually do in order to withstand the presence of H_2O_2 (Avery and Morgan, 1924).

How pneumococcal metabolism affects pneumococcal virulence, is an ongoing area of investigation. Mutations or removal of genes associated with key metabolic components have been shown to effect pneumococcal growth and survival (Gaspar *et al.*, 2014, Yesilkaya *et al.*, 2009, Tettelin *et al.*, 2001a). Additionally, the by-products of metabolism such as H_2O_2 have been implicated in altered virulence profiles (Pericone *et al.*, 2000, Loose *et al.*, 2015).

1.1.4 Identification of *Streptococcus pneumoniae*

The pneumococcus is closely related to other members of the *Streptococcus* genus in particular *S. mitis* and *S.oralis* both of which have been implicated in the horizontal transfer of homologous genes to *S.pneumoniae* (Dowson *et al.*, 1993, Teng *et al.*, 2002). Growth on blood agar produces distinctively alpha-haemolytic colonies due to the presence of H_2O_2. The extent of haemolysis helps to differentiate the pneumococcus from some of the other Streptococcal species, such as *S.pyogenes* which are beta-haemolytic on blood agar (Grahn *et al.*, 1983).

Differentiation from other alpha-haemolytic streptococci, e.g. *S.viridans*, may be achieved by testing the organism's bile solubility. The pneumococcus dissolves when immersed in bile salts due to the presence of the autolysin enzyme LytA which triggers an autolytic process (Mosser and Tomasz, 1970). Further differentiation may be achieved by assessing optochin sensitivity. The test is performed by placing an optochin (ethylhydrocupreine hydrochloride) disc on blood agar that has been streaked with a presumptive pneumococcal sample. Most pneumococci are sensitive to optochin which is evidenced by a zone of inhibition surrounding the optochin disc although resistance to optochin while not common has been observed (Pikis *et al.*, 2001).

1.1.5 Components of the pneumococcus

The majority of pneumococci are enveloped by a polysaccharide capsule which is recognised as the organism's prime virulence factor. Although there is significant diversity of capsular types they perform the same functional role namely to provide protection against phagocytosis (Hyams *et al.*, 2010) and mucus-mediated clearance (Nelson *et al.*, 2007).

In addition to the capsule, proteins found on the pneumococcal surface have inherent virulence potential as they may interact directly with host cells. They can be divided into four groups: lipoproteins; LPxTG (peptidoglycan binding motif); choline-binding proteins and; non-classical surface proteins (NCSPs) (Pérez-Dorado *et al.*, 2012).

Whilst pneumococci display remarkable interspecies diversity, several virulence factors have been commonly observed, e.g. the 'hair' like protein pilus and the cytoplasmic enzyme pneumolysin (Jedrzejas, 2001). Both components are an important part of pneumococcal virulence and are discussed alongside the surface proteins below. Figure 1 shows a diagrammatic representation of these virulence factors and where they may be found within the structure of a pneumococcus.

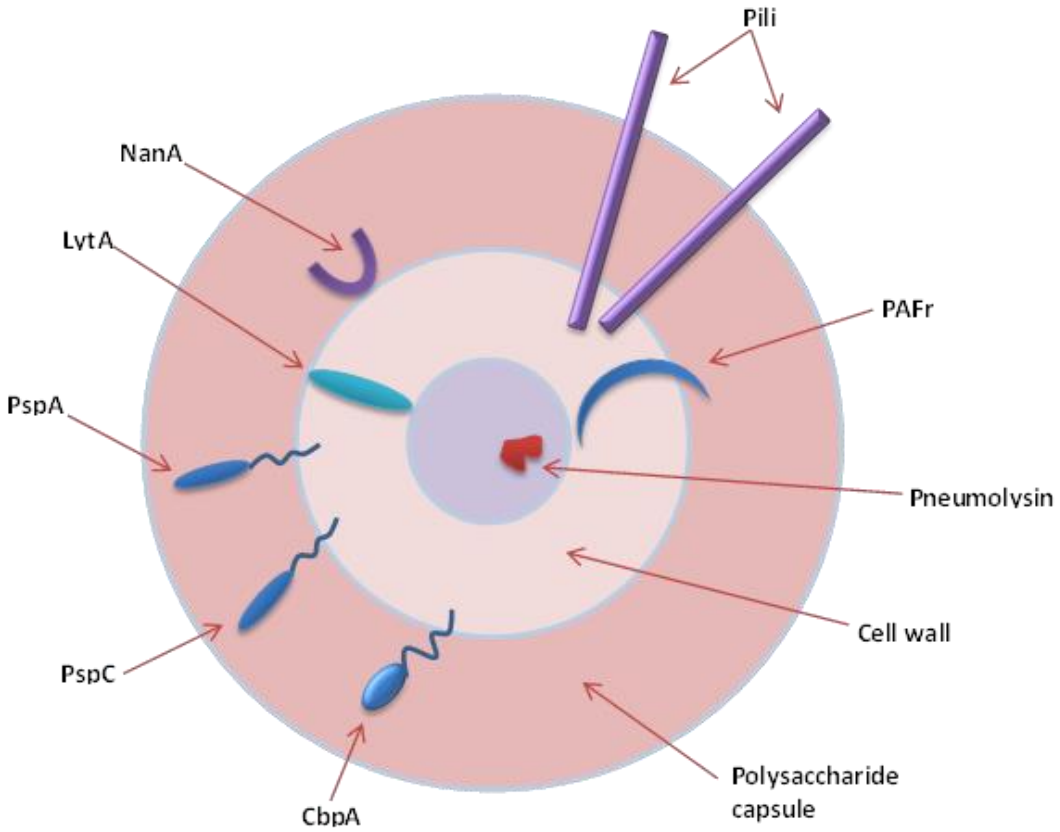


Figure 1 Structure of the pneumococcus.

1.1.5.1 Capsule

The first documented observation of the pneumococcal capsule was in 1881 when Pasteur described an 'aureole' or halo surrounding the organism when viewed under a light microscope (Pasteur, 1881). Recognition of its role in disease led to serological investigations and the discovery that pneumococci existed as multiple groups that could elicit distinct serological responses (Bezançon and Griffon, 1897, Eyre and Washbourn, 1899).

Investigations led by Neufeld in the early 1900s were able to demonstrate the agglutinating effect that pneumococci have when in contact with corresponding antisera. This work led to the differentiation of pneumococci into four groups (Neufeld, 1902, Neufeld, 1910). The swelling or 'quellung' reaction that he described is still a widely used pneumococcal typing method today.

Recognition of the immunological response to distinct capsular types gave rise to the term serotype. Over the following decades investigations into the possible use of antisera as a treatment for pneumonia revealed many other serotypes (Cooper

et al., 1932). By 1932, 32 serotypes had been described although the successful treatment of disease using antisera was limited.

The death of a Danish prince in 1939 expedited research into serotype diversity and the accuracy of typing methods (Vammen, 1939). Prince Vladimir of Denmark was treated with antisera against serotypes 9N and 9L however the treatment was unsuccessful. Further investigations found that a novel serotype belonging to serogroup 9 was the causative agent of his pneumonia. It was named 9V after the prince (Geno *et al*., 2015).

The Danish typing method which is used throughout the world today separates pneumococci into serogroups and serotypes. A serotype is defined as having a unique capsular structure. Serotypes that share many serological traits can be grouped together within a serogroup. To date 46 serogroups and over 90 serotypes have been identified (Bentley *et al*., 2006, Bratcher *et al*., 2010, Calix and Nahm, 2010, Calix *et al*., 2012, Oliver *et al*., 2013b, Geno *et al*., 2017, Park *et al*., 2015) although a number of recently discovered capsular types have been described as genetic variants of previously identified serotypes (Kapatai *et al*., 2017a, Mostowy *et al*., 2017).

The polysaccharide nature of the pneumococcal capsule was first described almost a century ago (Heidelberger and Avery, 1923). Since then, elucidation of the chemical composition has shown that each serotype has a unique polysaccharide structure composed of linkages, sugars and side chains with some arrangements leading to greater immunogenicity than others (Lammers *et al*., 2011). Synthesis of the capsule follows one of two processes: the synthase-dependent or the Wzy-dependent process. Only the capsules of two serotypes, 3 and 37 are synthesised via the synthase dependent process (Cartee *et al*., 2000, Llull *et al*., 2001) and all other serotypes follow the Wzy-dependent process (Yother, 2011). Genes responsible for capsule biosynthesis are clustered together on the capsular polysaccharide (*cps*) locus (Dillard *et al*., 1995) which sits between two flanking genes; *dexB* and *aliA*. Serotype 37 is the only exception having one gene, *tts*, associated with capsular biosynthesis, located in a different genomic region (Llull *et al*., 2001).

Capsular production can be expressed in a phase variable manner (Kim and Weiser, 1998). Variation in the opacity of colonies represents the amount of capsule polysaccharide produced with more opaque colonies having higher levels of capsule while transparent colonies have lower levels of capsule but higher

levels of teichoic acid (Saluja and Weiser, 1995). Observations of phase variation within mouse and rat models have demonstrated that these variations correlate with changes to virulence (Saluja and Weiser, 1995, Kim and Weiser, 1998, Manso *et al.*, 2014).

The main role of the capsule is to ensure pneumococcal survival by the avoidance of host defences and it has several mechanisms to achieve this. All but seven serotypes have capsules with a net negative charge (Nelson *et al.*, 2007). This is due to the acidic sugars, pyruvate, or phosphate present in the pneumococcal capsule. Of the remaining seven serotypes, six have a neutral charge (serotypes 7A, 7F, 14, 33F, 33A, and 37) while serotype 1 is zwitterionic (Kamerling, 2000). It has been proposed that the surface charge contributes to the avoidance of opsonphagocytosis through the electrostatic repulsion of host phagocytic cells (Li *et al.*, 2013).

Additionally, presence of a capsule restricts the access to surface antigens targeted by host defenses. Immunoglobulins and C-reactive protein play an important role in the activation of complement however physical interaction with the pneumococcus is impeded by the capsule (Suresh *et al.*, 2006, Hyams *et al.*, 2010). Interestingly, while the inhibitory effect of the capsule has many advantages to pneumococcal survival, its presence at times, is disadvantageous as it may also interfere with the attachment to host cells (Bergmann and Hammerschmidt, 2006). A mouse model of infection was able to demonstrate that non-capsulated pneumococci were able to colonize as successfully as capsulated strains (Park *et al.*, 2012) thus providing evidence that while the capsule is of great importance it is not the only determinant of pneumococcal survival.

1.1.5.2 The cell wall and associated proteins

The pneumococcal cell wall consists of teichoic acids, lipoteichoic acids, peptidoglycan, phophorycholine and a phospholipid membrane (Bergmann and Hammerschmidt, 2006).

Teichoic acid is the main component and plays an important role in host cell adherence. Platelet Activating Factor receptors (PAFr) are expressed by many mammalian cells and can bind with pneumococci via interactions with teichoic acid (Cundell *et al.*, 1995). Once bound, pneumococci may then be internalised by the host cell and escorted across epithelial barriers such as the blood-brain barrier or lung epithelium (Ring *et al.*, 1998).

Phosphorycholine (Pcho) is an important component of the cell wall. It is able to bind to the host PAFr and C-reactive protein thereby functioning as a pneumococcal adhesin. Additionally it serves as an anchor to choline-binding proteins (CBP) which includes cell wall hydrolases, the autolysins LytA, LytB and LytC and a phosphorylcholine esterase (Bergmann and Hammerschmidt, 2006). LytA has a particularly important role in pneumococcal virulence. Its activation leads to lysis of the cell and the release of the intracellular protein pneumolysin. The release of pneumolysin then promotes an inflammatory response which can contribute to exacerbation of disease in the host (Tuomanen *et al.*, 1985).

Other CBPs include Pneumococcal surface protein A (PspA) and Pneumococcal surface protein C (PspC). PspA offers significant protection to the pneumococcal cell by binding to apolactoferrin which is bactericidal and an important part of host defense (Shaper *et al.*, 2004). PspC can be divided into two different classes; one contains a CBP while the other is attached to the cell wall via a LPxTG motif. Both inhibit the binding of Factor H which regulates the alternative complement pathway and thereby prevents clearance by opsonophagocytosis (Janulczyk *et al.*, 2000).

The LPxTG family of proteins are anchored to the cell wall's peptidoglycan and include the neuraminidases: NanA, NanB and NanC; zinc metalloprotease (ZmpC) and the protease IgA1. These proteins appear to play an important role in pneumococcal adhesion and colonisation (King *et al.*, 2004, Weiser *et al.*, 2003, Oggioni *et al.*, 2003)

1.1.5.3 Pneumolysin

Pneumolysin is an intracellular pore-forming protein and a major virulence factor of the pneumococcus (Jedrzejas, 2001). At sub-lytic levels it can successfully inhibit a number of host defences including: complement deposition; the beating of cilia; and phagocytosis (Tilley *et al.*, 2005, Nandoskar *et al.*, 1986). It is released from the cytosol following cell lysis where it can damage host cells and provoke an inflammatory response (Tuomanen *et al.*, 1985).

1.1.5.4 Pili

Bacterial adhesion is enhanced by the presence of pili which are long hair-like organelles that extend beyond the pneumococcal capsule. Their presence provides a competitive advantage however they are only found in around 30% of clinical isolates (Barocchi *et al.*, 2006).

1.2 Pneumococcal typing methods

Recognition of the diversity of pneumococci has necessitated the development of typing methods. Understanding which pneumococcal strains are circulating can inform public health policies and assist in the design of interventions such as vaccines. Whilst serotyping provides essential information its use in epidemiology is limited. Additional typing methods based on pneumococcal molecular content have been developed to provide information regarding strain evolution and relatedness.

1.2.1 Serotyping

The Quellung reaction developed by Neufeld was the first method used to type pneumococci (Neufeld, 1902) and remains the gold standard of capsular typing methods today (Satzke *et al.*, 2013). The premise of the reaction is that a physical change may be observed when specific antibodies bind to the epitope of the corresponding capsular antigens. For the Quellung reaction, this change is seen as a swelling of the capsule when viewed with a light microscope.

This principle was applied to other typing methods such as latex agglutination (Kirkman Jr *et al.*, 1970) and co-agglutination (Smart and Henrichsen, 1986). These allow the interaction between epitope and antibody to be viewed with the naked eye as agglutination on a microscope slide. Statens serum Institut Diagnostica, Copenhagen, Denmark used this method to develop their Pneumotest-latex test which improved the speed and ease with which pneumococci are typed (Jauneikaite *et al.*, 2015).

In more recent years, knowledge of serotype-specific capsular genes has enabled the development of genotypic methods (Bentley *et al.*, 2006). Polymerase chain reaction, (PCR) uses this information and is widely employed in many molecular laboratories. Using short strands of DNA that are specifically matched to target sequences (primers); the pneumococcus may be serotyped through the amplification of these serotype-specific gene targets. This method is reliant on the accuracy of the primers as well as the quality of the samples. False positive results are a possibility in the presence of closely related strains (Carvalho *et al.*, 2013).

Technological advancements and improvements in affordability have made whole genome sequencing more available to researchers. *In silico* analyses allow the entire capsular locus to be interrogated to accurately type the pneumococcus.

Further, this method allows for the identification of single nucleotide polymorphisms (SNPs) to help distinguish closely related serotypes (Kapatai *et al.*, 2016).

1.2.2 Multi locus enzyme electrophoresis (MLEE)

Multilocus enzyme electrophoresis (MLEE) characterises organisms by detecting non-synonymous polymorphisms within a gene locus. Mutations that have led to amino acid substitutions in gene-coding enzymes affect the electrostatic charge. This can then be visualised as differences in mobility when the enzyme is subjected to electrophoresis (Selander *et al.*, 1986). This method has been used to assess the genetic relatedness of pneumococcal isolates of the same serotype to reveal that isolates with the same capsular type can possess underlying genetic differences which can be used to cluster the population in novel ways distinct from capsular type (Hall *et al.*, 1996).

1.2.3 Pulse field gel electrophoresis (PFGE)

PFGE was developed as a method to subtype bacterial strains in order to improve epidemiological investigations. DNA is first fragmented by restriction enzymes and then subjected to gel electrophoresis. The direction of the current is periodically altered to allow better separation of DNA molecules that are then visualised as a pattern of bands. PFGE has been of particular use in investigations of pneumococcal antibiotic resistance where it can further subtype strains with the same MLEE profiles (Lefevre *et al.*, 1993). The pneumococcal molecular epidemiology network (PMEN), who carry out global surveillance on antibiotic resistant pneumococcal strains, utilised this method in an effort to subtype pneumococci, assess relatedness and track global trends (McGee *et al.*, 2001).

1.2.4 Multi locus sequence typing (MLST)

In an effort to improve the portability of molecular typing results between laboratories, multilocus sequence typing (MLST) was proposed (Maiden *et al.*, 1998). The method was based on the principle of MLEE but instead of analysing structural changes to the enzyme, it relies on fragments of DNA sequences from stable 'housekeeping' genes. For the pneumococcus, seven 405 – 624bp fragments of the following genes were selected: *aroE* (shikimate dehydrogenase); *gdh* (glucose-6-phosphate dehydrogenase); *gki* (glucose kinase); *recP* (transketolase); *spi* (signal peptidase I); *xpt* (xanthine phosphoribosyltransferase);

and *ddl* (D-alnine-D-alanine ligase) (Enright and Spratt, 1998). Each sequence variation for each housekeeping gene is given a unique, arbitrary, allelic number. The combination of the seven numbers equates to a sequence type (ST).

There are a number of advantages to using MLST as a typing tool. Firstly the portability of this method means that data can be shared on-line (https://pubmlst.org) allowing comparisons to be made between studies. Secondly, evolutionary relationships can be inferred by assessing the number of shared alleles in closely related strains. Application of a specific algorithm (based upon related sequence types or 'BURST') to assess genomic relatedness leads to a population snapshot with bacterial isolates clustered according to allelic similarity (Feil *et al.*, 2004).

1.3 Pneumococcal interactions with the host

The pneumococcus is an opportunistic pathogen able to act as both an asymptomatic commensal and an aetiological agent of serious disease. The prevalence with which it features in both carriage and disease gives it a unique epidemiology with far-reaching public health implications (Price *et al.*, 2017).

Understanding the impact of this organism on human health is not straight forward. Not only must the role of the pneumococcus in carriage and in disease be examined but also its effect on both the individual host and the wider society.

1.3.1 Pneumococcal carriage

Bacterial carriage is a distinct phase that occurs following the acquisition and establishment of an organism. In the case of pneumococci, the human nasopharynx is the normal habitat and it may occupy it alone or alongside other bacterial species such as *Haemophilus influenzae*; *Staphylococcus aureus*; *Moraxella catarrhalis* and *Neisseria meningitidis* (Bogaert *et al.*, 2004b). The incidence of interbacterial co-colonisation depends on the organisms involved. Co-colonisation of the pneumococcus and *H.influenzae* has been found to occur to a greater degree than expected (Jacoby *et al.*, 2007) particularly as the pneumococcus possesses mechanisms to negatively affect competing colonisers. The production of H_2O_2 inhibits the growth of other bacteria while the presence of neuraminidase has been shown to remove sialic acid from the capsules of *H.influenzae* and *N.meningitidis* thus making those species more susceptible to opsonophagocytosis (Regev-Yochay *et al.*, 2006, Shakhnovich *et al.*, 2002).

The serotype of the colonising pneumococci may influence co-colonisation frequencies. An increase in carriage of *S.aureus* was observed following the introduction of a vaccine that targeted specific pneumococcal serotypes. Further investigations revealed an antagonistic relationship between pneumococci with vaccine type serotypes and *S.aureus* (Bogaert *et al.*, 2004c). Serotype-specific co-colonisation has also been examined using a microarray to detect the presence of multiple pneumococcal serotypes in carriage and found that multiple carriage is probably underestimated using standard identification methods (Turner *et al.*, 2011).

Pneumococcal colonisation is transient with the duration likely to be dependent upon a number of characteristics possessed by the host, the colonising strain or both. Host factors affecting carriage rates include age; geographical location; community setting and immune status (Bogaert *et al.*, 2011).

Rate of carriage is highest in children aged under five years and can range from 10% to 90% (Huang *et al.*, 2009, Bogaert *et al.*, 2011) .This wide range is due to the variability in specific ages and settings of carriage study participants. Stratifying children aged under five years into sub-age groups has shown that children aged <6 months and >4 years are colonised less frequently than other young children (Bogaert *et al.*, 2011, Huang *et al.*, 2009). This distribution of carriage rates amongst the under 5 age group is likely due to a combination of factors including an immature immune system and the level of social interactions. Children within day care centres have been found to have higher rates of carriage (Bogaert *et al.*, 2001, Dunais *et al.*, 2003) as have children that share familial settings with siblings (Principi *et al.*, 1999).

The geographical location of study participants can also lead to disparity in recorded carriage rates. Data collated from studies undertaken in sub-Saharan Africa found that carriage rates can be as high as 93% in some circumstances (Bogaert *et al.*, 2011). In the UK, most studies have calculated the rate to fall between 30 and 50% (Tocheva *et al.*, 2011a, van Hoek *et al.*, 2014, Gladstone *et al.*, 2015, Devine *et al.*, 2017).

Risk factors, besides age and geography that have been observed to effect pneumococcal carriage rates include: exposure to cigarette smoking (Greenberg *et al.*, 2006) ; recent antibiotic use (Petrosillo *et al.*, 2002, Principi *et al.*, 1999) and infection with other respiratory pathogens (Huang *et al.*, 2009, Smith *et al.*, 1976). It is difficult to elucidate whether specific risk factors directly affect the

acquisition of pneumococci or whether they lead to alterations in the density or duration of pneumococcal carriage which in turn may affect pneumococcal transmission (Thors *et al.*, 2016a).

How pneumococcal carriage relates to disease is an ongoing area of investigation. Whether disease is most likely to occur following the point of acquisition or following an extended period of colonisation is not known. It is however widely accepted that pneumococcal carriage is a precursor to pneumococcal disease (Simell *et al.*, 2012) and therefore remains an important area of research.

1.3.2 The burden of pneumococcal disease

The availability of disease surveillance data varies across the globe. In countries with nationalised health systems and robust surveillance procedures, high quality data can be produced. However, the limitation of data from some developing countries results in an under-representation of these geographical regions which hinders the culmination of accurate global disease estimates.

Efforts have been made to overcome these difficulties and provide good global estimates to reflect the degree of morbidity and mortality associated with pneumococcal disease. In 1990 the World Bank commissioned the first Global Burden of Diseases study which began collecting data relating to public health throughout the world including an estimate of the burden of pneumococcal pneumonia (IHME, 2016). The study has since fallen under the auspices of the World Health Organisation, WHO, and has now been running for over two decades and currently over 1000 researchers collaborate on annual estimates.

The Global Burden of Diseases, Injuries and Risk factors study 2016 estimated that over one million people died from pneumococcal pneumonia in 2016, including 340,000 children aged under 5 years (GBD, 2016). The number of disability-adjusted life years (DALYs) caused by pneumococcal pneumonia for all ages was almost 50 million (GBD, 2016). A report produced by Save the children (http://www.savethechildren.org.uk) has projected that 735,000 children will die of pneumonia in 2030 if the current trend is held (Save the children, 2017).

In 2009 the WHO published a comprehensive analysis of the global burden of disease in young children caused by the pneumococcus. Their assessment included data relating to cases of IPD and non-invasive pneumococcal disease in addition to the pneumococcal pneumonia estimates. The report indicated that in 2000 there were around 14.5 million cases of pneumococcal disease in children

aged under five years which resulted in approximately 826,000 deaths (O'Brien *et al.*, 2009).

In the UK the burden of pneumococcal pneumonia can only be estimated as this is not a notifiable disease and many cases are treated in the absence of laboratory confirmation. However, cases of invasive pneumococcal disease (IPD) are notified to Public Health England and Health Protection Scotland allowing accurate national statistics to be compiled.

Data collated by Public Health England (PHE) has shown that the burden of IPD in England has significantly reduced following widespread use of a conjugate vaccine from 15.63 cases per 100,000 (average cases between June 2000 and July 2006) to 6.85 cases per 100,000 in 2013/14 (Waight *et al.*, 2015) . These findings are in line with other European countries, Australia and North America (Myint *et al.*, 2013). Findings further reveal that the burden of disease is carried primarily by the youngest and oldest members of society (Hausdorff *et al.*, 2005) and it is these age groups that are generally targeted by public health interventions such as vaccination programmes.

1.3.3 Host response to the pneumococcus

When examining the relationship between host and microbe is it important to remember that both have coevolved over millennia to live in an equilibrious state (Majcherczyk and Moreillon, 2004). Disease occurs when this state becomes unbalanced and the microbial organism is able to proliferate either in an area where it doesn't normally reside or to an extent that is harmful for the host. For protection, the host has a number of mechanisms to respond to such a threat.

The human host has two systems to detect potential pathogens: innate immunity and adaptive immunity. The innate immune system is present from birth and is able to recognise many generic regions of microbial pathogens (Koppe *et al.*, 2012). Recognition is mediated by a number of proteins including Toll-like receptors (TLRs) and NOD-like receptors (NLRs) and a range of other proteins capable of stimulating an inflammatory response. TLRs can be further divided into 10 proteins each with a different role in pathogen mediation. One of these, TLR-2 recognises components of the pneumococcal cell wall and along with TLR-9 is able to enhance pneumococcal phagocytosis. TLR-2 may also play an important role in pneumococcal translocation by weakening epithelial barriers and allowing pneumococci to proliferate elsewhere (Koppe *et al.*, 2012).

NLRs can also be subdivided and it is one of these proteins; NOD-2 that recruits macrophages to the area of invasion. In a mouse model of infection NOD-2 was shown to have a likely role in meningitis as mice lacking NOD-2 were observed to have less demyelination, inflammation and astrogliosis (Koppe *et al.*, 2012).

Additional to TLRs and NLRs the innate system has inflammasomes which are complexes of proteins capable of inducing the inflammatory cascade. Deficiencies within the inflammasome complex have been associated with increased susceptibility to pneumococcal pneumonia as they appear to have a role in bacterial clearance and the maintenance of epithelial and endothelial barriers (Koppe *et al.*, 2012).

The adaptive immune system is more specialised and based upon previous interactions with specific antigens. It is triggered when phagocytes digest invading pathogens and present small fragments of the invader via class 2 major histocompatibility molecules to helper T cells (T_H cells). The activated T_H can differentiate into T_H1 cells which activate macrophages or T_H2 cells which activate B cells to produce antibodies (Bogaert *et al.*, 2009).

The immune system takes a number of years to mature and this is exemplified by higher carriage and disease rates in young children (Hausdorff *et al.*, 2005, Bogaert *et al.*, 2004c). Children under 5 years of age are poor at generating antibodies against capsular polysaccharide antigens therefore interventions such as vaccines must use a protein conjugate to elicit a sufficient immune response (Adderson, 2001) .

Organs of the body which are involved in the immune system include the spleen, the thymus, lymph nodes and bone marrow. Any host conditions that effect any of the immune system organs either increases the likelihood of pneumococcal disease or worsens the prognosis of a diseased state (Delves and Roitt 2000).

1.3.4 Pneumococcal disease

Disease occurs when the pneumococcus is able to move to an area of the host that is not normally colonized, evade host defences and proliferate. The area to which the pneumococcus moves defines the type of disease it causes.

Pneumococcal presence in a normally sterile site results in a diagnosis of invasive pneumococcal disease (IPD) such as meningitis, bacteraemia or infections of the joints and soft tissue. Pneumonia is not classed as invasive unless a secondary

bacteraemic infection accompanies it. Other infections caused by the pneumococcus includes: acute otitis media (AOM); conjunctivitis and bronchitis.

Infection of the meninges with *S. pneumoniae* is responsible for the most severe form of meningitis leading to higher mortality rates in children aged <5 years than other bacterial organisms such as *N. meningitidis* and *H. influenzae* (Châtelet *et al.*, 2005). It is also more likely to cause seizures, coma and focal neurology than other forms of bacterial meningitis (Randle *et al.*, 2011). Early diagnosis of infants is often difficult as the disease can resemble many common viral infections. Older children typically present with the 'classic' symptoms of headache, fever and neck stiffness.

Pneumococcal bacteraemia may develop as a primary infection or following an episode of pneumonia. It is the most common bacterial pathogen associated with bacteraemia in children with no underlying risk factors (Gomez *et al.*, 2014) and more likely to occur in older age groups and the immunosuppressed (Feikin *et al.*, 2010).

The pneumococcus is also responsible for many non-invasive diseases such as AOM, sinusitis and conjunctivitis (Bluestone *et al.*, 1992, Gwaltney Jr *et al.*, 1994, Barker *et al.*, 1999). AOM occurs when pneumococci are carried into the eustachian tubes and often follows a viral infection that has caused congestion (Gray *et al.*, 1980b). Similarly, congestion of mucosal membranes within sinus cavities, either by infection or allergy can lead to sinusitis (Gwaltney Jr *et al.*, 1994). Interestingly, many pneumococci that have been isolated in cases of conjunctivitis have been unencapsulated and although unencapsulated pneumococci do feature in other disease states, their presence is unusual and observations of capsulated strains are more common (Hilty *et al.*, 2014). Whilst IPD is a major burden on global health it is outweighed by the incidence of non-invasive pneumococcal diseases (Feldman and Anderson, 2014).

Pneumococcal pneumonia has historically been acknowledged as the most common form of community acquired pneumonia (Heffron, 1939). It occurs when pneumococci are able to reach the alveoli and proliferate. The infection incites an inflammatory response leading to the build up of fluid and white blood cells in the alveolar space. Pneumonia frequently follows a viral infection and is particularly associated with the influenza virus (Kim *et al.*, 1996).

There are many underlying conditions that either increase the probability of contracting pneumonia or alter the host's ability to recover from it. Diseases

affecting the lungs such as asthma and chronic obstructive pulmonary disorder (COPD) have been shown to alter bacterial clearance leading to infections of greater duration and a higher chance of mortality (Crim *et al.*, 2009). Similarly cigarette smoking and alcohol abuse have been shown to the affect the host's ability to resist infection (Nuorti *et al.*, 2000, Perlino and Rimland, 1985). In the absence or failure of treatment, pneumonia can lead to pleural empyema, bacteraemia or ultimately death (Byington *et al.*, 2002, Marrie, 1992).

1.4 Clinical interventions

The early 1910s saw the development of a number of treatment strategies designed to reduce the burden of disease caused by the pneumococcus. This included experimentation with type-specific antiserum which followed recognition that the pneumococcus has a diverse range of capsular types each able to provoke a distinct serological response (Avery *et al.*, 1918, Horsfall *et al.*, 1937).

At around the same time, early antibiotics were being trialled in mouse models (Morgenroth and Levy, 1911) and the first pneumococcal vaccine was being developed (Butler *et al.*, 1999). It was these strategies that proved most advantageous to human health and the decades that followed have been built upon these early experiments.

1.4.1 Antibiotics and resistance

Optochin, a derivative of quinine was first tested in a mouse model of pneumococcal infection (Morgenroth and Levy, 1911). A year later, it was recorded that pneumococci were displaying resistance to it thus producing the first ever report of antibiotic resistance (Austrian, 1975). Despite this optochin was used for a short time to treat pneumonia and conjunctivitis until evidence emerged of eye damage caused by the toxicity of the drug as well as increasing levels of resistance (Moore and Chesney, 1917).

Sulfonamides, the first of the broad spectrum antibiotics, became available in the 1930's. Sulfanilamide, an early sulphonamide was used successfully to treat streptococcal infections however efficacy against pneumococcal infections was limited (Long and Bliss, 1937). A sulfanilmide congener, sulfapyridine produced better results and was used to some success in the management of

pneumococcal pneumonia (Evans and Gaisford, 1938) however again, resistance was soon to emerge (Lowell *et al.*, 1940).

The discovery of penicillin heralded a great change both in the successful treatment of pneumococcal infections and in the attitude of medical researchers. Serotyping and vaccine development became less important as antibiotics appeared to be the wonder drug the medical world had been waiting for (Kauffman, 1979). Resistance to penicillin was shown early *in vitro* (Schmidt and Sesler, 1943) however there were no significant cases of penicillin resistance in human infection so it continued to be the drug of choice. Pharmacotherapy options were expanded with the availability of cephalosporins (Murdoch *et al.*, 1964) and both drugs continued to be used to successfully treat pneumococcal disease for decades.

Clinically relevant resistance to penicillin was first reported in the 1960s (Hansman *et al.*, 1971, Hansman and Bullen, 1967). For many bacterial species, penicillin resistance is attributable to the proliferation of beta-lactamases which are usually encoded within a plasmid and transmitted between bacteria. Despite the naturally transformable nature of the pneumococcus, penicillin resistance is not conveyed in this manner but instead through structural changes within specific proteins (Hakenbeck *et al.*, 1999).

Pneumococci possess six penicillin binding proteins (PBPs): PBP1A, PBP1B, PBP2A, PBP2B, PBP2X and PBP3. Penicillin and other β-lactams bind to PBPs which inhibits cell wall synthesis (Appelbaum, 2002). Resistance arises when mutations occur in the genes encoding the PBPs. The level of resistance is dependent upon the gene in which the mutation occurs (Dowson *et al.*, 1994). Higher levels of resistance have been observed when PBP2B is involved (Barcus *et al.*, 1995).

Whilst point mutations within PBPs are the main mechanism for pneumococcal penicillin resistance, acquisition and incorporation of genomic fragments from other pneumococci or from other streptococcal species such as *S.mitis* and *S.oralis* (Sibold *et al.*, 1994) have been found to confer antibiotic resistance.

The development of nonpenicillin antimicrobials followed soon after penicillin. The macrolide antibiotic, erythromycin inhibits pneumococcal protein synthesis and has been successfully used to treat pneumococcal infection since the 1950s. Pneumococci that express the *mefE* gene are capable of efflux of macrolides from the pneumococcal cell and are therefore resistant. Additionally, pneumococci that possess *erm* methylases confer resistance through the modification of rRNA

which is a macrolide target site. Surveillance of antibiotic resistance in the USA revealed that macrolide resistance doubled between 1995 and 1999 from 10% to 20% of invasive pneumococcal isolates (Hyde *et al.*, 2001).

Tetracycline and chloramphenicol were also developed shortly after penicillin and have also been used to treat pneumococcal infection for many decades (Finland *et al.*, 1976). They invoke a bacteriostatic effect on bacterial growth by binding to subunits of the ribosome and inhibiting protein synthesis (Kohanski *et al.*, 2010). As with erythromycin, resistance is acquired through the horizontal transfer of genes from other pneumococci or streptococcal species. The *tet* and *cat* genes are responsible for tetrycline and chloramphenicol resistance respectively and the majority of them are found on integrative conjugative elements (ICEs); *tet* is associated with Tn*912* and *cat* is usually found on Tn*5252* (Croucher *et al.*, 2009a, Wyres *et al.*, 2013). Asymmetrical co-transfer of *tet* and *cat* resistance genes has been identified with tetracycline resistance almost always found with chloramphenicol resistance but the reverse, chloramphenicol resistance with associated tetracycline resistance, is found to a lesser extent (Shoemaker *et al.*, 1979).

Resistance has been reported for every class of antimicrobial used to treat pneumococcal infection (Klugman, 1990). A newer antibiotic, vancomycin was first used in 1981 (Garau *et al.*, 1981) and is considered a last resort, therefore its use is protected. However a number of cases of vancomycin tolerance have been reported both in vitro and in clinical isolates (Moscoso *et al.*, 2010).

In the late 1970's and early 1980's, pneumococci displaying resistance to several antibiotics were observed (Jacobs *et al.*, 1978, Radetsky *et al.*, 1981). The Pneumococcal Molecular Epidemiology Network (PMEN) was set up in 1997 to undertake global surveillance of pneumococcal clones with resistance to multiple antibiotics (Klugman, 1998). The first three PMEN clones were thought to originate in Spain and were typed as: 23F, 6B and 9V. More recently the network has decided to include some antibiotic susceptible clones that have high invasiveness and a wide global spread (PMEN website, 2014).

Recognition of the extent of pneumococcal antibiotic resistance led to renewed interest in pneumococcal typing and the development of vaccines to target strains observed to be clinically relevant either due to their level of invasiveness or to association with antibiotic resistance (Austrian, 1980).

1.4.2 Pneumococcal vaccines

In 1911, while attempting to limit disease in South African gold miners, Sir Almroth Wright began development of a vaccine made from heat killed pneumococci. A reduction in infection rates was seen in the months following immunisation, however, the overall morbidity rates remained unchanged (Butler *et al.*, 1999). In the decades that followed, the composition and variability of the pneumococcal capsule was further scrutinised leading to the development of more targeted vaccines.

The widespread use of penicillin in the 1940s heralded a change in the control of pneumococcal disease. Daily administration of benzylpenicillin saw a huge decrease in the incidence of disease cases making vaccination seemingly unnecessary (Tillett *et al.*, 1944). This was later re-evaluated when the mortality rates of high risk populations were observed to be increasing (Austrian and Gold, 1964).

In the decades that followed attention turned back to vaccine development however this time with an increased understanding of both pneumococcal heterogeneity and host immunology.

1.4.2.1 Polysaccharide vaccines

Polyvalent polysaccharide vaccines were proven to be successful in lowering morbidity and mortality in numerous trials. The introduction first of a 14-valent, then a 23-valent vaccine to the US saw a decline in disease rates, however, due to the large number of component parts of a 23-valent vaccine, the aggregate efficacy is less than those of vaccines targeting a smaller number of infections (Austrian, 1999). This slight lowering of efficacy means that repeated exposure to a variety of serotypes puts the vaccine recipient at a statistical risk of developing disease. Additionally, the use of polyvalent polysaccharide vaccines, whilst displaying favourable results in adult populations, was observed to be less effective for infants (Sloyer *et al.*, 1981). As mentioned previously, the thymus-independent polysaccharide capsule does not provoke a sufficient immune response from immature immune systems.

1.4.2.2 Conjugate vaccines

When a protein is chemically linked to a pneumococcal polysaccharide capsule the resultant conjugate vaccine is proven to elicit a far greater immune response than polysaccharide alone (Goebel *et al.*, 1932).

In 2000, Prevenar™, a heptavalent conjugate vaccine developed by Wyeth (now part of Pfizer) was licensed for use in Europe and the US. It was added to the UK's national immunisation programme in 2006. The vaccine contains capsular antigens of seven of the most prevalent serotypes found in disease: Serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. These are conjugated to a highly immunogenic protein, the non-toxic diphtheria toxin variant carrier protein, CRM_{197} which increases the efficacy of the vaccine by eliciting a T-cell dependent immune response (Pletz *et al.*, 2008).

In 2010 following surveillance of circulating serotypes post-PCV7, a tridecavalent pneumococcal conjugate vaccine, PCV13, superseded PCV7 in the UK PCV13 contains antigens from the original seven serotypes included in PCV7 plus an additional six serotypes: 1, 3, 5, 6A, 7F and 19A.

An alternative ten-valent vaccine, PCV10 made by GlaxoSmithKline (GSK, 2017) was licensed at a similar time to PCV13. Its formulation includes antigens from ten serotypes, eight of which: 4, 6B, 9V, 14, 23F, 1, 5 and 7F are conjugated to non-typeable *Haemophilus influenzae* protein D. Of the remaining two serotypes, 19F is conjugated to a diphtheria toxoid while 18C is conjugated to a tetanus toxoid (Vesikari *et al.*, 2009). The composition of each of the vaccines is shown in Table 2.

By 2016, 135 countries had introduced PCV into routine immunisation programmes (IVAC report, 2016). Of these, 93 countries (69%), including the UK, were using PCV13, 30 countries (22%) were using PCV10 and 11 countries (8%) were using both. Studies comparing the impact of PCV10 and PCV13 on IPD rates have found no significant difference between the overall efficiay of each vaccine (IVAC report, 2016).

There is some variation in the scheduling of vaccine doses. Of the 135 countries using PCV, 108 (80%) currently have a 3 dose schedule. Of these, 50 countries offer a 2 + 1 schedule meaning 2 primary doses are given and one booster when the infant is aged 12 – 15 months, and 48 countries offer a 3 + 0 schedule meaning 3 primary doses are given, usually when an infant is aged 2, 3 and 6

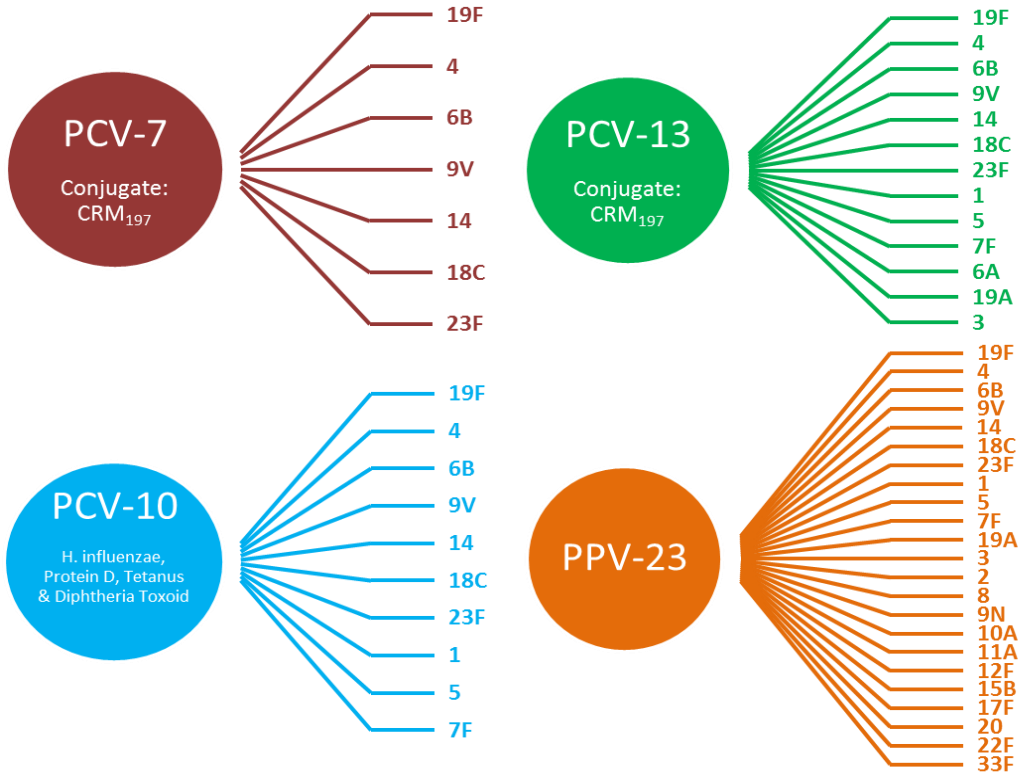
months with no booster (Whitney *et al.*, 2014, IVAC report, 2016). Scheduling of PCV doses is an area of constant review. In October 2017 the Joint Committee on Vaccination and Immunisation (JCVI) recommended that the UK switch to a 1 + 1 schedule however, at the time of writing this schedule was still being discussed within the scientific community (O'Brien, 2017).

Table 2 PCV vaccination schedule, UK

Recommended age	Vaccine
8 weeks	First dose PCV
16 weeks	Second dose PCV
12 to 13 months	Booster dose PCV

Recommended schedule at time of writing (June 2018).

Figure 2 Pneumococcal vaccine composition



1.4.2.3 Serotype replacement

PCV7 was licensed for use in 2000. The seven serotypes selected for the vaccine's formulation were not only highly burdensome in disease but also in carriage. Removal of seven prevalent serotypes would leave a large ecological niche and it was unknown how other serotypes or other bacterial species would respond to this. It was predicted that the niche would be filled by other serotypes not covered by the vaccine (Lipsitch, 1997) and/or that vaccine type (VT) serotypes would undergo genomic changes in order to evade removal (Spratt and Greenwood, 2000).

Following the widespread usage of PCV7 there was an observed decline in the incidence of both disease and carriage featuring any of the included serotypes (Tocheva *et al.*, 2011a, Gladstone *et al.*, 2015, Devine *et al.*, 2017, Waight *et al.*, 2015). Surveillance studies indicated that this decline was echoed in unvaccinated populations which demonstrated the herd immunity effect conferred by the widespread implementation of PCVs (Miller *et al.*, 2011, Davis *et al.*, 2013).

Whilst the rate of vaccine type (VT) serotypes decreased significantly following vaccine introduction, an increase was seen in non-vaccine type (NVT) serotypes (Gladstone *et al.*, 2015). The six additional serotypes included in PCV13 were responsible for 69% of IPD cases post-PCV7 and this formed the basis of their inclusion in the updated PCV13 (Ladhani *et al.*, 2013).

Since PCV13's introduction in 2010, Public Health England (PHE) has reported a 32% decrease in IPD caused by VT serotypes compared to a pre-PCV13 baseline (Ladhani *et al.*, 2013). The overall decrease has however been offset by serotype replacement with NVT serotypes (Waight *et al.*, 2015). A report from PHE for the epidemiological year 2013/14 showed that NVT serotypes were the most common cause of IPD. Serotype 8 was the most prevalent in two age groups: 5-64 years and 65 years and over and serotypes 24F, 15B/C and 22F were the most prevalent in children aged under 5 years.

NVT serotypes were also observed to dominate post-PCV carriage. Prevalence of serotypes 19A, 11A, 15B, 21, 22F, 6C and 23B were reported following PCV7 introduction (Tocheva *et al.*, 2011b, Flasche *et al.*, 2011) and all of these continued to dominate carriage in children following PCV13 introduction (Devine *et al.*, 2017, van Hoek *et al.*, 2014) with the exception of 19A which was included in the updated vaccine. Serotype 15A was observed to increase in children aged four and under in the three years post PCV13 (Devine *et al.*, 2017) while serotypes 24F and 6C were the most prevalent serotypes found in children and young adults aged 5 to 20 years in the two years post PCV13 (van Hoek *et al.*, 2014). Fluctuations in serotype prevalence have been observed (Devine *et al.*, 2017, Gladstone *et al.*, 2015) therefore the long-term impact post PCV13 serotype distribution requires continued surveillance of pneumococcal carriage to compare with serotype-specific disease incidence.

1.5 Pneumococcal genomics

Recognition of the pneumococci's underlying genetic material can be dated from Griffiths' famous experiment in 1928. Using live rough (unencapsulated) pneumococci and heatkilled smooth (encapsulated) pneumococci, to inoculate mice, Griffiths was able to demonstrate the 'transforming principle' by recovering live smooth bacteria from the mice thus demonstrating that information is exchanged between strains (Griffith, 1928). In 1944, Avery, MacLeod and McCarty took this further when their experiments established that it was deoxyribonucleic acid, DNA that is exchanged (Avery *et al.*, 1944).

Interest in DNA has continued ever since and led to such landmark moments as Watson and Crick's discovery of DNA structure (Watson and Crick, 1953) and Sanger's advancements in genomic sequencing (Sanger *et al.*, 1977b). Recently the development of computational tools with which to analyse genomic sequences have opened up new avenues for researchers to explore.

1.5.1 First generation sequencing technologies

Nucleic acid sequencing is the process for determining the exact sequence of the nucleic acids bases, Adenine, Thymine (in DNA only, Uracil in RNA only), Cytosine and Guanine within a molecular target. The earliest attempts at sequencing were primarily focussed on short RNA fragments which involved the separation of radiolabelled nucleotides. These early attempts were able to reveal the composition of nucleotides within a protein gene but not their order (Holley *et al.*, 1961).

Adaptations to these methods led to the whole DNA genome of the bacteriophage ΦX174 being sequenced in 1977 (Sanger *et al.*, 1977a). Sanger used radiolabelled nucleotides which attached to a primer using DNA polymerase. DNA synthesis occurred from a primer template and continued until the labelled terminator was incorporated into a strand. Fragments from each reaction could then be separated using electrophoresis and the resultant bands indicated the sequence of each of the bases (Heather and Chain, 2016).

This was the beginnings of 'first generation sequencing technology' which was spurred on by another breakthrough of Sanger; the use of dideoxynucleotides (ddNTPs). ddNTPs are monomers of a DNA strand that lack a 3′ hydroxyl group therefore incorporation of one into an extending DNA chain terminates the process (Chidgeavadze *et al.*, 1984). The method became known as the dideoxy chain-termination method (often referred to as Sanger sequencing) and it originally used four parallel reactions each with a different ddNTP base to infer a nucleotide sequence. Later developments involving the use of fluorolabelling instead of radiolabelling, reduced the need for four parallel reactions and allowed the process to take place in one vessel (Ansorge *et al.*, 1986). This allowed the process to become automated which opened up sequencing possibilities to other researchers (Hunkapiller *et al.*, 1991).

Further developments such as the use of shotgun sequencing (Anderson, 1981) allowed longer fragments to be sequenced while the use of polymerase chain reaction (PCR) (Saiki *et al.*, 1988) made DNA targets more easily available.

1.5.2 Second generation sequencing technologies

Second generation sequencing technologies (SGSTs) moved away from the first generation by using light emissions to indicate the sequence of nucleotides and by scaling up the number of sequences they were able to produce in parallel. A number of different automated sequencing methods were developed which differed in terms of cost, speed and accuracy.

Pyrosequencing (Margulies *et al.*, 2005), sequencing-by-ligation (McKernan *et al.*, 2009) and ion semiconductiong sequencing (Rothberg *et al.*, 2011) are three methods that have been harnessed by companies to develop sequencing platforms each with their own strengths and drawbacks. Pyrosequencing is used in 454 sequencing™ (now owned by Roche) which was the first high throughput sequencing machine. Sequencing-by-ligation was utilised in the SOLiD™ (Sequencing by Oligonucleotide Ligation and Determination) sequencer, and the ion semiconducting sequencing was performed on the Ion Torrent™ .A brief overview of each of these platforms is given in Table 3.

Sequencing Platform	Method
454 sequencing™ Life Sciences (Roche)	DNA fragments are captured on beads and amplified via emulsion PCR. Beads are then washed over a plate that contains wells designed to fit one bead per well. dNTPs and the enzyme luciferase are washed over the beads and light is emitted by the luciferase when a dNTP is incorporated.
SOLiD sequencing™ Applied Biosystems (Life technologies)	DNA fragments are captured on beads and amplified via emulsion PCR. Beads are then bound to a glass slide before fluorescently-labelled probes and DNA ligase are added. Synthesis occurs via a series of ligation steps and the assignation of a nucleotide is based upon which probe ligated.
Ion Torrent™ sequencing (Life technologies)	A DNA template is added to a microwell. A single type of nucleotide is added and, if incorporated, releases a detectable amount of hydrogen ions. The more nucleotides incorporated, the greater the number of ions released.

Table 3: Overview of some SGSTs

1.5.2.1 Illumina sequencing

Illumina sequencing has dominated the SGST era and become the most widely used sequencing technology in genomic research (Quail *et al.*, 2012). A number of platforms are available from benchtop sequencers such as the Miseq and the Nextseq series to production-scale sequencers such as the Hiseq and the Novaseq series. Whilst there are differences in how the platforms are applied they each share the same basic process.

Illumina sequencing uses a bridge amplification method whereby fragments of purified DNA are ligated to adaptors and loaded onto a solid surface flow cell. The surface of the cell is covered in complementary oligonucleotides for the adaptors to bind to. The addition of nucleotides and polymerisation enzymes leads to amplification of the ligated strands. Each cycle of denaturation separates

the strands and creates more templates for amplification. This process results in clusters of clonally amplified DNA strands which can then be sequenced.

Primers, polymerase and four labelled dNTP terminators are added to the templates on the flow cell. Identification of bases occurs once a laser identifies fluorescence emitted from the labelled dNTP. This process is repeated for each of the bases on the DNA fragment.

Illumina is a sequencing-by-synthesis approach that has a number of advantages over competitors. The paired-end data it produces is read first from one end of cell-bound DNA then, following an extension cycle and reversal of the DNA strand relative to the flow-cell, it performs a second read from the opposite end of the first read. This assists with assembling the reads post-sequencing. The read length and depth of each of the Illumina platforms varies for example the HiSeq 2500 is capable of 4 billion reads per run and produces paired end reads of around 125bp whereas the MiSeq has a lower throughput at 25 million sequencing reads per run however it produces longer read lengths (2x250bp).

1.5.3 Third generation sequencing technologies

The use of the term third generation sequencing technologies is contentious as there is not yet a clear definition of what constitutes the third generation. However a number of advancements have been made which separate newer sequencing methods from their predecessors.

The PacBio series (Pacific Biosciences) are single molecule real time (SMRT) platforms able to produce extremely long reads (~10kb) in real time (van Dijk *et al.*, 2014). The process uses tiny holes called zero-mode waveguides (ZMWs) in which DNA polymerisation occurs (Levene *et al.*, 2003). Light passes through the ZMWs to illuminate the bottom of a well and detect DNA polymerase molecules. DNA chains are extended by washing over the DNA library with fluorescent dNTPs. The long reads produced are particularly useful for de novo assembly of genomes.

Nanopore sequencing has been utilised by Oxford Nanopore Technologies (ONT) who offer a number of platforms including the GridION and the PromethION for larger scale projects and the highly portable, pocket-sized, MinION. It is the portability of the MinION sequencer that has captured many researchers' interests and it has been successfully used in extreme environments such as Antarctica (Johnson *et al.*, 2017) and even aboard the International Space Station (Castro-

Wallace *et al.*, 2016). ONT's newest sequencer, in development at the time of writing, the SmidgION promises to increase portability further by utilising smartphone technology.

Nanopore DNA sequencing occurs when molecules are driven through a nano-scale hole (a nanopore) which is set in an electrically resistant membrane. A voltage is run through the membrane and an ionic current passes through the nanopore. Nucleotides that pass through the pore interrupt the current in a way that is characteristic of the specific nucleotide (Haque *et al.*, 2013). Accuracy has improved since the MinIONs launch in 2014 due to the availability of updates to the sequencer, the flow cells and the sequencing kits (Jain *et al.*, 2017)

1.5.4 The genome of the pneumococcus

The genome of the pneumococcus didn't become publically available until 2001 when three genomes were published within a few months of each other. The first was a draft genome of G54, a serotype 19F clinical isolate (Dopazo *et al.*, 2001). The other two were the complete genomes of TIGR4, a virulent serotype 4 isolate (Tettelin *et al.*, 2001b) and R6, an avirulent laboratory strain with a 7,504 bp deletion that removes the *cps* locus from the genomic sequence (Hoskins *et al.*, 2001). The R6 was a descendant of the rough strain used in Avery's experiments of the 1940s labelled D39, and this was sequenced six years after R6 (Lanie *et al.*, 2007). The genome of D39 contains the *cps* locus that is deleted in R6 (serotype 2) as well as the plasmid pDP1.

Analyses of TIGR4 and R6 revealed that these pneumococcal genomes are over 2Mb long, contain over 2,000 protein coding sequences (CDSs) and have a relatively low GC content of 39.7% (Tettelin and Hollingshead, 2004). TIGR4, R6 and a further 13 pneumococcal genomes surveyed by Tettelin and Hollingshead revealed that pneumococci possess a variety of insertion sequences (IS) which provides evidence of considerable genomic plasticity and adaption to a variety of ecological niches. Most of the repeated sequences can be divided into BOX and RUP (Repeat Unit of Pneumococcus) elements although further analyses on one of the PMEN clones (ATCC 700669) have revealed another repeat family which has been named SPRITE (*Streptococcus pneumoniae* Rho-Independent Terminator-like Elements) (Croucher *et al.*, 2011b).

BOX elements are variable arrangements of repeat sequences named boxA, boxB and boxC. Their presence can either stimulate or inhibit the expression of

downstream genes depending on their genomic position (Knutsen *et al.*, 2006). RUP elements are simple 107bp repeat units that are frequently found inserted into or near insertion (IS) elements. It has been hypothesized that their presence limits the functionality of transposase genes within the genome (Delihas, 2008).

Pneumococcal genomes also contain a relatively high proportion of genes dedicated to the catabolism of sugars. Reliance on sugar transport to provide the organism's energy has led to a large number of transporters mostly of the ATP-binding cassette (ABC) or phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) types. This system of ATP and PEP generation while inefficient allows an advantage over other respiratory tract colonizers such as *H.influenzae* and *N.meningitidis* which generally have a far more limited variety of transporters.

A number of methods have been employed to deduce the function of genes found within the pneumococcal genome including the development of microarrays. Although microarrays had been available in some form since the 1970s (Grunstein and Hogness, 1975) the availability of whole genome sequences and oligonucleotides allowed the development of oligo arrays which increased the specificity of genomic target regions (Bumgarner, 2013). Microarrays have been particularly useful in the study of gene expression and have uncovered a number of genes associated with virulence such as those coding for the major autolysin (*lytA*), pneumolysin (*ply*) and the pyruvate oxidase (*spxB*) (Orihuela *et al.*, 2004). The use of microarray-based comparative genomic hybridization (CGH) has further revealed the pathogenicity island *psrP* which has been shown to increase the invasiveness of different pneumococcal serotypes (Obert *et al.*, 2006, Blomberg *et al.*, 2009).

Whilst microarrays have been useful in eliciting information from genomic data, the improved cost-effectiveness of whole genome sequencing has led to an increase in using large-scale sequencing of multiple isolates to compare bacterial populations. This approach can be used to identify genetic determinants of colonisation or disease (Li *et al.*, 2015); investigate diversity within a species (Croucher *et al.*, 2014); and infer evolutionary relationships (Chewapreecha *et al.*, 2014).

The genomic content of all isolates within a given population has been termed the pan-genome which comprises of a core genome, i.e. those genes shared by all isolates of a population, and an accessory (or dispensable) genome, i.e. genes

that vary between isolates. A model to estimate the size of the pneumococcal pan-genome predicted that it would continue to grow as more genomes are sequenced because novel genes would continue to be revealed (Donati *et al.*, 2010).

A study to identify the core genome of the pneumococcus predicted that 397 genes were essential for fitness (Van Opijnen *et al.*, 2009) however further work which examined over 3000 isolates collected from four locations: Southampton (UK); Boston (USA); Maela (Thailand) and Reykjavik (Iceland) revealed that only 127 of these genes were found in all isolates (van Tonder *et al.*, 2017). This study highlighted the heterogeneity of pneumococci from different global regions.

Whilst extensive variation has been reported in the gene content of different bacterial populations (Baumdicker *et al.*, 2012, Collins and Higgs, 2012), a recent analysis of pneumococcal accessory genomes has found a surprising similarity in the frequency at which accessory genes are distributed (Corander *et al.*, 2017). Furthermore the pattern of distribution appeared to be irrespective of specific lineages and unaffected by clinical interventions such as the widespread use of conjugate vaccines.

1.5.5 Mechanisms for gene exchange

Bacteria use three main mechanisms to transfer DNA: transformation; transduction and conjugation.

Transformation is the uptake of exogenous DNA and the pneumococcus has a dedicated competence system for this process. Exogenous DNA is taken into the pneumococcal cell via a pseudopilus before it is degraded and incorporated into the pneumococcal genome (Pestova and Morrison, 1998). Competence is initiated by an extracellular signal (Tomasz and Mosser, 1966) which activates a Competence Stimulating Peptide (CSP). CSP detection alters the physiology of the pneumococcal cell and this change is referred to as the X-state (Claverys *et al.*, 2006). Whilst in the X-state, pneumococci are able to lyse nearby cells to release DNA (Claverys *et al.*, 2007). As levels of amino acids (aa) increase, the X-state becomes repressed (Claverys *et al.*, 2000).

Whilst fragments of incorporated DNA are generally around 4.4kb, much larger fragments have been observed with one study reporting transformation involving an entire *cps* locus along with its flanking PBP genes (Brueggemann *et al.*, 2007).

Transduction is the phage-mediated transfer of DNA and a number of pneumophages have been identified (Ramirez *et al.*, 1999, Romero *et al.*, 2009). Choline binding proteins (CBP) are thought to act as phage receptors and therefore expression of a capsule can be inhibitory (Bernheimer and Tiraby, 1976). A correlation between prophage acquisition and adherence to eukaryotic cells has been reported (Loeffler and Fischetti, 2006) so it is possible that their presence increases the chance of pneumococcal colonisation.

Conjugation occurs when two cells are in direct contact with each other and are able to transfer mobile elements such as plasmids or integrative and conjugative elements (ICEs). The pneumococcus has been associated with two types of plasmid: pDP1 (Smith and Guild, 1979) and pSpnP1 (Romero *et al.*, 2007).

ICEs include conjugative transposons and integrative plasmids which are both transferred in a circular form before integrating via site-specific recombination (Burrus *et al.*, 2002). They are frequently associated with antibiotic resistance. The WGS of *S.pneumoniae* ATCC 700699, a serotype 23F, ST81 clone with observed resistance to multiple antibiotics, revealed an ICE which included a Tn*916*-like element inserted into a Tn*5252*-like transposon. This arrangement is common to streptococcal species and is important as it confers resistance to both tetracycline and chloramphenicol (Croucher *et al.*, 2009b).

1.5.6 Bioinformatic analyses

Numerous methods for performing genomic analyses exist with a wide variety of tools for performing similar tasks. Despite this a general consensus is emerging with many studies following a similar approach. Generally, for epidemiological investigations, isolates that have undergone whole genome sequencing are aligned and mapped against a reference to identify single nucleotide polymorphisms (SNPs). A statistical method may then be applied to the alignment to infer phylogeny. As the pneumococcus has been shown to be highly recombinogenic (Croucher *et al.*, 2011a), many studies first attempt to limit the confounding effect of recombination by distinguishing regions of dense polymorphisms and removing them from the phylogeny thus leaving an evolutionary tree based upon vertical inheritance.

Phylogenomic investigations have been used to examine the emergence and diversification of pneumococcal lineages following PCV introduction. This is a particularly useful approach for understanding changes in antibiotic resistance.

Whilst PCV usage has led to an overall reduction of antibiotic resistant pneumococcal disease, the incidence of disease caused by antibiotic resistant NVT strains has increased (Gertz Jr *et al.*, 2010). Several studies have traced the evolution of such strains and found that they existed prior to vaccine introduction and their increased prevalence was likely to fill the niche left when VT strains were reduced by PCVs (Andam *et al.*, 2017, Croucher *et al.*, 2011a). Understanding how pneumococci evolve in response to vaccine intervention through examination of populations prior and post PCV may help to predict future changes and assess the impact of further clinical interventions (Gladstone *et al.*, 2017).

In addition to phylogenomic analyses, the pneumococcal genome can be interrogated to identify non-capsular genes of interest. This is particularly useful for investigating non-capsulated pneumococci. Associated more frequently with carriage than disease, non-capsulated pneumococci have been observed to have an extensive array of accessory genes and may act as a reservoir of virulent genes that can be shared with capsulated pneumococci (Chewapreecha *et al.*, 2014).

Comparisons between invasive and non-invasive pneumococci can further our understanding of genetic determinants of disease; may help to explain the success of some pneumococcal strains over others and; provide useful information regarding future vaccine design (Li *et al.*, 2015, Cleary *et al.*, 2016).

Whilst this is undoubtedly an area of increasing value, it is still a developing discipline with no single approach able to be applied. Pipelines do exist, and generalised algorithms are relatively simple to apply, however these are not always fully understood or used to their most advantageous. For the majority of researchers the full potential of what bioinformatics can reveal is yet to be discovered.

1.6 Pneumococcal transmission

Pneumococcal survival is dependant not only on successful colonisation but also on transmission from host to host in order to proliferate. The ecological niche in which the organism abides is an ideal base as it provides opportunity for spread via aerosolised droplets.

The rate of pneumococcal colonisation differs with age with young children having the highest rates of pneumococcal carriage (Gray *et al.*, 1980a, Smith *et al.*, 1993). It is thought that the nasopharynx of young children is the major

reservoir for pneumococci going on to cause serious invasive disease in older adults (Hussain *et al.*, 2005, Melegaro *et al.*, 2004) and that different serotypes and clonal types have different transmission characteristics (Melegaro *et al.*, 2007).

Evidence of pneumococcal transmission from children to older adults has been provided through studies within familial settings (Dowling *et al.*, 1971, Owen Hendley *et al.*, 1975) and from seasonal disease rates (Walter *et al.*, 2009). However, not all studies have shown proof of transmission from children to older adults (Regev-Yochay *et al.*, 2004) which demonstrates that the dynamics of this process is likely to be complex being not only dependent upon the host's age but also on the capsular type of transmitted pneumococci (Weinberger *et al.*, 2016).

The availability of WGS data has been useful in the investigation of transmission due to the high resolution with which it can explore multiple genomes. Exploration of transmission can be divided into three scales: direct host to host; local transmission and; global transmission.

Investigation of host to host transmission is particularly useful for examining the origin of outbreaks and to distinguish nosocomial infections from community acquired infections. Bacteria are isolated from the dense sampling of a defined host population so bacterial genomes may be compared to identify variance. The number of SNPs that differ between isolates will indicate the likelihood of transmission. The threshold, of how many single nucleotide polymorphisms (SNPs) indicate transmission, is set according the bacteria of interest. This method has been used to investigate *Clostridium difficile* in hospital patients (Eyre *et al.*, 2013) and *Mycobacterium abscessus* in cystic fibrosis patients (Bryant *et al.*, 2013).

Local transmission is carried out by sampling a subsection of a population to identify transmission within a defined area which could be as small as a hospital or as large as a country. Transmission is indirect with unsampled individuals making up part of the chain (Jombart *et al.*, 2014). This approach clusters isolates based on genomic similarities and can reveal genuine disease outbreaks as well as patterns of cryptic transmission where the source is not immediately apparent. Studies utilising this approach have been used to investigate the spread of methicillin-resistant *Staphylococcus aureus* (MRSA) in a neonatal unit and the surrounding community (Harris *et al.*, 2013) as well as an outbreak of *Escherichia coli* O104:H4 (Rasko *et al.*, 2011).

Global transmission is used to trace the spread of microorganisms across the world. Methods for investigating global spread generally rely on the construction of phylogenomic trees to infer transmission routes. This approach has been used to investigate a foot and mouth disease outbreak (Ypma *et al*., 2013) . The use of WGS to track bacterial pathogens has also been used to identify transmission between populations of people for example the spread of *Neisseria gonorrhoeae* from a homosexual population into a heterosexual population (Grad *et al*., 2014).

1.7 Study rationale

The burden of disease caused by the pneumococcus led to the inclusion of PCV7 into the UK's NIP in 2006. Studies published since have shown that while the pneumococcal carriage rate has remained stable the incidence of pneumococcal disease has decreased (Gladstone *et al.*, 2015, Miller *et al.*, 2011). Pneumococcal carriage has undergone a rapid shift regarding serotype distribution caused by the replacement of vaccine type (VT) serotypes with non-vaccine type (NVT) serotypes.

In 2010, PCV7 was replaced by PCV13 which included six additional serotypes. The effect of this replacement on carriage and disease requires ongoing surveillance to ensure continued vaccine effectiveness against IPD. Whilst some data has been published (Waight *et al.*, 2015) it has been shown that the pneumococcus is highly adaptable (Devine *et al.*, 2017) and therefore unlikely to reach equilibrium for a number of years following vaccine introduction.

The data presented within this thesis is derived from pneumococci isolated from carriage and cases of IPD collected over 10 years. The carriage isolates are from children aged 4 years and under and were collected each winter from 2006/07 to 2015/16. The IPD isolates and associated data are from the same timeframe and were collected from the Public Health England (PHE) laboratory located at University Hospital Southampton NHS Foundation Trust.

The analysis of pneumococcal data collected over 10 years during which time PCV13 was introduced provides an opportunity to study vaccine impact on both carriage and disease. Further by comparing data from two populations (carriage and disease) during a time when the bacterial population is under selective pressure from PCV13, insights may be gained regarding herd immunity and pneumococcal transmission.

The availability of whole genome sequencing has provided an opportunity to study molecular changes that occur within a population of isolates and compare them with another population. Application of these methods will be utilised to assess if any changes occurring in carriage are also occurring in IPD. Through understanding similarities and differences between the two populations it is hoped that the impact of PCV13 will be more fully understood.

1.8 Hypothesis

Population changes seen in pneumococci isolated from asymptomatic carriage during the implementation of PCV13 will also be seen in pneumococci isolated from invasive pneumococcal disease matched for location and time.

1.8.1 Aims

- To define how the introduction of PCV13 has impacted the population of carriage isolates between 2006 and 2016
- To define how the introduction of PCV13 has impacted the population of IPD isolates between 2006 and 2016
- To examine if changes in the carriage population are also observed in the IPD population.

1.8.2 Objectives

- To characterise the population of pneumococci isolated from paediatric carriage over a 10 year period through evaluation of carriage rates and serotype distribution.
- To characterise the population of pneumococci isolated from cases of IPD over a 10 year period in terms of incidence rates, serotype distribution and age specific differences.
- To describe and compare the molecular epidemiology of carriage and disease isolates.
- To compare the impact of PCV13 on carriage and IPD.

Chapter 2 Methods

2.1 Study statement and contributors

The paediatric pneumococcal carriage study is an on-going study that began in winter 2006/07. Three former PhD students Dr Anna Tocheva; Dr Rebecca Gladstone and Dr Vanessa Devine were responsible for serotype and MLST results prior to 2013/14. Their contributions are listed as part of my declaration at the beginning of this thesis. A summary of their analysis methods is provided in Table 1.

2.1.1 Ethics

Ethical approval for this study was granted by Southampton and South West Hampshire Committee 'B' (REC numbers 05/Q1704/105 and 14/NS/1064).

Written and verbal information relating to the study was provided to the parent/guardian of each study participant to ensure informed consent. Written consent was obtained before study participation.

2.1.2 Study outline

The paediatric pneumococcal carriage study commenced in 2006, its main aim was to assess epidemiological changes to the pneumococcal population following the introduction of a seven-valent pneumococcal conjugate vaccine (PCV7) to the UK's national immunisation programme (NIP). The study takes nasopharyngeal swabs from children aged four years and under attending the outpatients department of Southampton General Hospital. Participants were excluded if outside the age criteria or if they were participating in other medical studies. Only one member per family was swabbed to reduce bias.

In study years one to seven, nasopharyngeal swabbing was performed by trained medical students. In years eight to ten, research nurses from the Clinical Research Facility (CRF) performed all nasopharyngeal swabbing.

In study year five (winter 2010/11), a questionnaire was introduced for the parent/guardian to complete. This provided information relating to the participants: age; vaccine status; recent respiratory health and antibiotic usage. The questionnaire was updated in 2013/14 to include influenza and meningitis B

vaccination status. After obtaining a research passport, I was able to accompany the research nurses and assist in the recruitment of study participants.

2.1.3 Study samples

Nasopharyngeal samples were obtained using a rayon tipped, pernasal, medical wire (TRANSWAB®, Medical Wire & Equipment, Corsham, UK.) and inserted into Amies media with charcoal. For the first six years of the study, swabs were sent to the on-site laboratory of the Health Protection Agency, HPA, (now Public Health England) for processing as per their standard operating protocol. From year seven (2012/13), swabs were processed within the laboratory of the research group and overseen by a registered Biomedical Scientist (BMS).

As a BMS registered with the Health and Care Professions Council (HCPC) (Registration number: BS66664), I was competent to carry out the processing of NP swabs received in years eight to ten.

2.1.4 Sample size

For the first eight years of the study, the sample size was set at 100 pneumococcal isolates per swabbing period. The number of isolates was determined by using a power analysis to estimate the number of isolates that would need to be collected in order to detect a change in carriage. The lowest expected carriage rate was estimated to be 10%. In order to detect a 50% change in this estimated carriage rate, a sample size of ≥100 pneumococcal positive isolates would allow us a 5% significance level with 80% power.

From the ninth study year, following REC approval, swabbing continued until the end of the swabbing period (March) taking the number of pneumococcal isolates beyond the 100 collected in previous years.

2.2 Microbiological processing

For years one to six, microbiological processing occurred in the laboratory of the HPA (now PHE). Swabs were processed within nine hours of collection and plated onto Columbia Colistin Naladixic Acid agar (CNA, Oxoid, Basingstoke, UK) and incubated for 18hrs at 37˚C in 5% CO_2. Presumptive pneumococci were streaked onto Columbia Blood agar (CBA, Oxoid) with an optochin disc (Oxoid). Isolates that displayed optochin sensitivity were recorded as positive and ten individual colonies were sub-cultured then stored at -80 ˚C on cryogenic beads

(Microbank™, Pro-Lab Diagnostics, UK). For the first two study years, only presumptive pneumococci underwent further investigation and storage. From year three, other selective media were utilised in order to detect the presence of additional bacterial species. Table 4 details the inclusion of additional species, and the year that they were included in the study.

From year seven, swabs were processed in our own laboratory within six hours of collection. Selective media were used in order to detect species of interest. For pneumococcal detection, all nasopharyngeal swabs were streaked onto CNA media and an optochin disc was placed onto the agar to check for optochin susceptibility. Plates were then placed in a 5% CO_2 incubator and grown overnight at 37˚C. Colonies that displayed pneumococcal morphology with a >14mm inhibition zone surrounding the optochin disc were recorded as positive and purity plated in preparation for storage. Swabs of pure pneumococcal isolates were inserted into 1mL of a liquid medium containing skim milk, tryptone, glucose and glycerine (STGG) and stored at -80˚C.

Table 4 Inclusion of bacterial species

Study Year	Winter	Presence of bacteria cultured for in study samples[1]
1	2006/07	S.pn
2	2007/08	S.pn
3	2008/09	S.pn; H.inf; MSSA; MRSA; AHS; N.men
4	2009/10	S.pn; H.inf; MSSA; MRSA; AHS; N.men; M.cat
5	2010/11	S.pn; H.inf; MSSA; MRSA; AHS; N.men; M.cat
6	2011/12	S.pn; H.inf; MSSA; MRSA; AHS; N.men; M.cat
7	2012/13	S.pn; H.inf; MSSA; MRSA; AHS; N.men; M.cat
8	2013/14	S.pn; H.inf; MSSA; MRSA;AHS; N.men; M.cat
9	2014/15	S.pn; H.inf; MSSA; MRSA; AHS; N.men; M.cat
10	2015/16	S.pn; H.inf; MSSA; MRSA; AHS; N.men; M.cat

The pneumococcus is the focus of the study however isolates of other species have been identified and collected from the same nasopharyngeal swab. Many isolates form the basis of other people's research.

[1] S.pn= *Streptococcus pneumoniae*; H.inf= *Haemophilus influenzae*; MSSA = *methicillin sensitive Staphylococcus aureus*; MRSA= *methicillin resistant Staphylococcus aureus*; AHS= *alpha-haemolytic streptococcus*; N.men= *Neisseria meningitides*; M.cat= *Moraxella catarrhalis*

2.2.1 IPD isolate collection

Pneumococci isolated from laboratory confirmed cases of IPD were stored on microbeads (Microbank™, Pro-Lab Diagnostics, U.K) and kept at -80°C until samples were ready for DNA extraction. Pneumococci isolated from blood, CSF and normally sterile sites were deemed causative of an invasive infection. Metadata relating to all IPD isolates were sent each month to a facilitator within our research group. Isolates were matched to the metadata in 2014/15.

2.3 DNA extraction

DNA was extracted from pneumococcal isolates in preparation for whole genome sequencing which was undertaken by staff at the Wellcome Trust Sanger Institute (WTSI).

2.3.1 Growth of pneumococcal cultures

Isolates were plated from the stored STGG samples onto CBA (Oxoid) plates and incubated overnight at 37°C in 5% CO_2. A single colony was then sub-cultured onto half a CBA plate and incubated under the same conditions overnight. Providing the resultant growth appeared pure, a sweep of colonies would then be used to inoculate 3mL of Brain Heart Infusion (BHI) broth. The same swab would be used to streak half a plate of CBA as a purity check before commencement of DNA extraction. If purity plates showed mixed growth the corresponding BHI broth would be deemed contaminated and the process would be repeated from the original stored STGG.

2.3.2 Pre-lysis and DNA extraction

Approximately 1.2mL of the prepared BHI broths was aliquoted into sterile eppendorf tubes and centrifuged at 12,500 rpm for 3 minutes. Supernatant could then be poured off to leave a pellet of cells in the bottom of each eppendorf.

Presence of the pneumococcal capsule makes it necessary to perform a pre-lysis step to ensure successful DNA extraction and this was achieved by using freshly prepared lysozyme (10mg/mL of lysozyme was added to nuclease-free water). 200µL of lysozyme was used to resuspend the bacterial pellet before being incubated at 37˚C for one hour.

Following pre-lysis of pneumococcal isolates, DNA extraction was performed using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany. Cat no. 51306). 50µL of the Qiagen AL lysis buffer and 20µL of Proteinase K was added to each sample. Each tube was then inverted to mix the suspension and incubated at 56˚C for one hour. A further 150µL of lysis buffer AL was added to each sample and the tubes were mixed by inversion and left at room temperature for ten minutes. Finally, 250µL of 100% molecular grade ethanol was added to each sample and mixed by inversion before transferring the entire suspension to a Qiagen spin column ready for DNA washing and elution.

Qiagen columns were placed in collection tubes and spun twice at 8000rpm for one minute; the first spin with the transferred suspension and the second spin with 500µL of buffer AW1. New collection tubes were used for each spin. 500µL of buffer AW2 was added to each column and spun at 14,000rpm for three minutes before the columns were transferred to elution tubes. I used an elution buffer that fulfilled the recommendations of The Sanger Institute, a Tris pH8 EDTA buffer (TE Buffer, Thermo Fisher Scientific). 60µL of pre-warmed (56°C) buffer was added to each column, incubated at room temperature for five minutes and spun at 12,000rpm for two minutes. Finally the columns were discarded and the eppendorf tubes, with 60µL of eluted DNA, were retained.

2.3.3 DNA quantification

The Qubit® fluorometer (Thermo Fisher Scientific, Waltham, US) was used with a dsDNA broad range (BR) assay kit (Thermo Fisher Scientific) which is capable of measuring concentrations of 2 – 1000ng. A working solution was prepared at room temperature using a 199µL of BR buffer to 1µL of light sensitive BR reagent for each sample to be tested. The Qubit® was calibrated using two standards by adding 10µL of each standard to separate qubit tubes containing 190µL of the prepared working solution. The two standards were then measured prior to the extracted DNA samples. For the DNA samples, 195µL of the working solution was aliquoted into a qubit tube and 5µL of extracted DNA was added. The Qubit® fluorometer could then read the concentrations after the correct sample volume was selected via an on-screen prompt. The Sanger Institute request a minimum sample DNA concentration of 20ng/µL in a total volume of 50µL.

2.3.4 Storage and shipment

Samples were stored at -20˚C until ready for overnight shipment on dry ice to the WTSI. Samples were aliquoted into a 96 well plate (ABgene AB-0800, Thermo Fisher Scientific) with the last well (H12) left empty. Each plate was marked with clear identifiers and sealed with a foil sealer. An accompanying metadata sheet was included in the shipment and e-mailed to the WTSI ahead of shipment.

2.4 Whole genome sequencing

Whole genome sequencing (WGS) was performed entirely by the WTSI using Illumina HiSeq 2000 AND 2500 instruments. Approximately 100 *S. pneumoniae* isolates from every year of study were sequenced, quality checked, assembled and annotated. This resulted in multiple files per isolate: FASTQ files containing raw sequencing data; FASTA files containing the assembled genomes and General Feature Format (GFF) files containing annotations of each genome. All files were then compressed into GNU zip files and sent via File Transfer Protocol (FTP) to be stored and analysed on the University of Southampton's High Performance Computing facility, Iridis.

WGS data pertaining to study years 1 – 5 were received and analysed by Dr Rebecca Gladstone as part of her PhD. Years 6 and 7 were received and analysed by Dr Vanessa Devine as part of her PhD and years 8 to 10 were received and analysed by myself.

FASTQ files of all carriage isolates from all years were stored in the European Nucleotide Archive (www.ebi.ac.uk) which is publically accessible. WGS analysis for which I required genomic files for years prior to 2013/14 was achieved following retrieval of files via the ENA.

2.4.1 Assembly

Whole genomes were assembled at the WTSI via their HPC facility. Velvet (European Bioinformatics Institute, UK) was used for the *de novo* assembly of short sequences to produce FASTA files. Files were then compressed and sent via FTP where they were stored, unzipped and analysed on Iridis.

2.4.2 Quality control

The Sanger Institute performed post assembly quality checks on each pneumococcal isolate. Results of these checks were contained within a comma-separated values (csv) file and sent via FTP. This automated process compares the sequencing data to a reference genome, in this case *S. pneumoniae* ATCC 700669, and a percentage of the genome that is identically mapped is produced. The csv file also contains information such as: depth of coverage; number of contigs; total length and N50 (a statistical measurement of length). An overview of what these quality measurements relate to is listed in Table 5.

Quality checks were also carried out on receipt of FASTQ files using the online, publically available software FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) which generates a report following submission of paired FASTQ files.

Table 5 Description of sequencing statistics

Sequencing Statistic	Description
Depth of coverage	The average read length is divided by the expected length of the genome (ref genome is ATCC700669 with a length of 2,221,315bp) multiplied by the number of reads.
Number of contigs	Contigs are contiguous sequences of data assembled from shorter reads. A smaller number of contigs indicates a less fragmented assembly.
Total length	Total length is an indication of quality as the length of the assembly shouldn't be too dissimilar from the reference genome.
% mapped	All isolates were mapped against the reference genome, ATCC700669 to see what percentage was able to align.
N50	A statistical measurement that is roughly similar to the median of the length of the genome. It represents the length that all contigs of ≥length contain at least half of the size of all contig lengths that make up the assembly. A smaller N50 indicates increased fragmentation.

2.5 Pneumococcal typing

Several methods were employed to characterise the pneumococcal isolates. Methods used in earlier study years have been described elsewhere (Tocheva *et al.*, 2011a, Gladstone *et al.*, 2015, Devine *et al.*, 2017). In study years 8 to 10 and for all IPD isolates, determination of the pneumococcal capsular type was deduced using whole genome data as well as traditional phenotyping methods. Whole genome sequences were also used to determine sequence type (MLST).

2.5.1 Slide agglutination

A Neufeld antisera kit was purchased from The Statens Serum Institut (Copenhagen, Denmark) and used to determine capsular type via slide agglutination. I followed the standard operating procedure (SOP) that is used by Public Health England. Pneumococcal isolates stored in STGG at -80˚C were briefly thawed and a sterile loop was used to inoculate half of a CBA plate. After overnight growth at 37˚C in a 5% CO_2 incubator, a single colony was purity picked, spread onto another half CBA plate and re-incubated overnight. After inspecting the plates for signs of mixed growth, a single loopful of pure growth pneumococci was transferred to 5mL of Todd Hewittt Broth and re-incubated.

After 12 hours of incubation, the broths were spun at 1500rpm for 30 minutes to form a bacterial pellet at the bottom. Supernatant was poured off until~200µl remained in the tube. The samples were then vortexed to re-suspend the pellet.

10µl of suspension was pipetted onto a glass slide and a loop of selected antisera was added and mixed into the suspension. The slide could then be rocked for up to 30 seconds and observed for signs of agglutination. If agglutination was observed, a positive result was recorded. If no agglutination was observed, a fresh 10µl of suspension was tested against another antisera. Antisera could be classified as pool, group or type. Pool antisera contain different serogroups and form the first level of agglutination investigations (Table 6). Through a process of elimination, a positive result obtained from pool antisera will direct the analysis to the set of group antisera that constitute the second level of investigation (Table 7). Once a positive result is obtained from one of the group antisera, type specific antisera can then be used until an ultimate identification of a serotype is reached.

If no agglutination was achieved with any of the pool antisera, then an omniserum containing antibodies to 91 serotypes was used to test for capsular

presence. If a negative result was achieved with the omniserum, then further testing of the isolate was carried out using WGS data.

Table 6 Pool antisera and associated groups

Pool Anitsera	**Serogroups**
Pool A	1, 2, 4, 5, 18
Pool B	3, 6, 8, 19
Pool C	7, 20, 24, 31, 40
Pool D	9, 11, 16, 36, 37
Pool E	10, 12, 21, 33, 39
Pool F	17, 22, 27, 32, 41
Pool G	29, 34, 35, 42, 47
Pool H	13, 14, 15, 23, 28
Pool I	25, 38, 43, 44, 45, 46, 48
Pool P	1, 7, 14, 19
Pool Q	6, 18, 23
Pool R	3, 4, 9, 12
Pool S	5, 8, 10, 15, 17
Pool T	2, 11, 20, 22, 33

Table 7 Group antisera and associated serotypes

Group Antisera	**Serotype**	**Pool**
6	6A, 6B, 6C,6D	B, Q
7	7F, 7A, 7B, 7C	C, P
9	9A, 9L, 9N, 9V	D, R
10	10F, 10A, 10B, 10C	E, S
11	11F, 11A, 11B, 11C, 11D	D, T
12	12F, 12A, 12B	E, R
15	15F, 15A, 15B, 15C	H, S
16	16F, 16A	D
17	17F, 17A	F, S
18	18F, 18A, 18B, 18C	A, Q
19	19F, 19A, 19B, 19C	B, P
22	22F, 22A	F, T
23	23F, 23A, 23B	H, Q
24	24F, 24A, 24B	C
25	25F, 25A	I
28	28F, 28A	H
32	32F, 32A	F
33	33F, 33A, 33B, 33C, 33D	E, T
35	35F, 35A, 35B, 35C	G
41	41F, 41A	F
47	47F, 47A	G

2.5.2 *In silico* capsular typing

Two methods of *in silico* capsular typing were used to characterise the pneumococcal isolates from study years 8 and 9.

SRST2 (Inouye *et al.*, 2014) was used by mapping the FASTQ files of each pneumococcal isolate against a file that contained the concatenated sequences of 91 pneumococcal serotypes. 90 of the pneumococcal serotypes were published by Dr Stephen Bentley (Bentley *et al.*, 2006) and serotype 6C was also included in my analysis (Elberse *et al.*, 2011). Reads that mapped >90% were reported. If more than one serotype was reported then the isolate would undergo further investigation. This included a more detailed look at the capsular sequence and/or follow up with slide agglutination analysis.

An *in silico* simulation of a PCR was also used to identify capsular type. Ipcress (Slater and Birney, 2005) uses an input file containing pairs of primers to identify different capsular types. WGS in FASTA format is used and sequences that are found within the specified primers are searched for. The primer files used were recommended by the Centre for Disease Control and Prevention, Atlanta, Georgia, USA (CDC) and are shown in Table 8.

Table 8 Primers for investigation of capsular type

Serotype	Gene	Primer Sequence (5' – 3')	Size (bp)
1	*wzy*	Fwd CTCTATAGAATGGAGTATATAAACTATGGTTA Rv CCAAAGAAAATACTAACATTATCACAATATTGGC	280
2	*wzy*	Fwd TATCCCAGTTCAATATTTCTCCACTACACC Rv ACACAAAATATAGGCAGAGAGAGACTACT	290
3	*galU*	Fwd ATGGTGTGATTTCTCCTAGATTGGAAAGTAG Rv CTTCTCCAATTGCTTACCAAGTGCAATAACG	371
4	*wzy*	Fwd CTGTTACTTGTTCTGGACTCTCGATAATTGG Rv GCCCACTCCTGTTAAAATCCTACCCGCATTG	430
5	*wzy*	Fwd ATACCTACACAACTTCTGATTATGCCTTTGTG Rv GCTCGATAAACATAATCAATATTTGAAAAAGTATG	362
6A/6B/6C/ 6D	*wciP*	Fwd AATTTGTATTTTATTCATGCCTATATCTGG Rv TTAGCGGAGATAATTTAAAATGATGACTA	250
6C/6D	*wciN*□	Fwd CATTTTAGTGAAGTTGGCGGTGGAGTT Rv AGCTTCGAAGCCCATACTCTTCAATTA	727

7C/7B/40	*wcwL*	Fwd CTATCTCAGTCATCTATTGTTAAAGTTTACGACGGGA Rv GAACATAGATGTTGAGACATCTTTTGTAATTTC	260
7F/7A	*wzy*	Fwd TCCAAACTATTACAGTGGGAATTACGG Rv ATAGGAATTGAGATTGCCAAAGCGAC	599
8	*wzy*	Fwd GAAGAAACGAAACTGTCAGAGCATTTACAT Rv CTATAGATACTAGTAGAGCTGTTCTAGTCT	201
9N/9L	*wzx*	Fwd GAACTGAATAAGTCAGATTTAATCAGC Rv ACCAAGATCTGACGGGCTAATCAAT	516
9V/9A	*wzy*	Fwd GGGTTCAAAGTCAGACAGTGAATCTTAA Rv CCATGAATGAAATCAACATTGTCAGTAGC	816
10A	*wcrG*	Fwd GGTGTAGATTTACCATTAGTGTCGGCAGAC Rv GAATTTCTTCTTTAAGATTCGGATATTTCTC	628
10F/10C/33C	*wzx*	Fwd GGAGTTTATCGGTAGTGCTCATTTTAGCA Rv CTAACAAATTCGCAACACGAGGCAACA	248
11A/11D	*wzy*	Fwd GGACATGTTCAGGTGATTTCCCAATATAGTG Rv GATTATGAGTGTAATTTATTCCAACTTCTCCC	463
12F/12A/44/ 46	*wzx*	Fwd GCAACAAACGGCGTGAAAGTAGTTG Rv CAAGATGAATATCACTACCAATAACAAAAC	376
13	*wzx*	Fwd TACTAAGGTAATCTCTGGAAATCGAAAGG Rv CTCATGCATTTTATTAACCGCTTTTTGTTC	655
14	*wzy*	Fwd GAAATGTTACTTGGCGCAGGTGTCAGAATT Rv GCCAATACTTCTTAGTCTCTCAGATGAAT	189
15A/15F	*wzy*	Fwd ATTAGTACAGCTGCTGGAATATCTCTTC Rv GATCTAGTGAACGTACTATTCCAAAC	434
15B/15C	*wzy*	Fwd TTGGAATTTTTTAATTAGTGGCTTACCTA Rv CATCCGCTTATTAATTGAAGTAATCTGAACC	496
16F	*wzy*	Fwd GAATTTTTCAGGCGTGGGTGTTAAAAG Rv CAGCATATAGCACCGCTAAGCAAATA	717
17F	*wciP*	Fwd TTCGTGATGATAATTCCAATGATCAAACAAGAG Rv GATGTAACAAATTTGTAGCGACTAAGGTCTGC	693
18A/18B/18C/ 18F	*wzy*	Fwd CTTAATAGCTCTCATTATTCTTTTTTTAAGCC Rv TTATCTGTAAACCATATCAGCATCTGAAAC	573
19A	*wzy*	Fwd GAGAGATTCATAATCTTGCACTTAGCCA Rv CATAATAGCTACAAATGACTCATCGCC	566
19F	*wzy*	Fwd GTTAAGATTGCTGATCGATTAATTGATATCC Rv GTAATATGTCTTTAGGGCGTTTATGGCGATAG	304
20	*wciL*	Fwd GAGCAAGAGTTTTTCACCTGACAGCGAGAAG Rv CTAAATTCCTGTAATTTAGCTAAAACTCTTATC	514
21	*wzx*	Fwd CTATGGTTATTTCAACTCAATCGTCACC Rv GGCAAACTCAGACATAGTATAGCATAG	192
22F/22A	*wcwV*	Fwd GAGTATAGCCAGATTATGGCAGTTTTATTGTC Rv CTCCAGCACTTGCGCTGGAAACAACAGACAAC	643

23A	*wzy*	Fwd TATTCTAGCAAGTGACGAAGATGCG Rv CCAACATGCTTAAAAACGCTGCTTTAC	722
23B	*wzx*	Fwd CCACAATTAGCGCTATATTCATTCAATCG Rv GTCCACGCTGAATAAAATGAAGCTCCG	199
23F	*wzy*	Fwd GTAACAGTTGCTGTAGAGGGAATTGGCTTTTC Rv CACAACACCTAACACTCGATGGCTATATGATTC	384
24A/24B/24F	*wzy*	Fwd GCTCCCTGCTATTGTAATCTTTAAAGAG Rv GTGTCTTTTATTGACTTTATCATAGGTCGG	99
31	*wzy*	Fwd GGAAGTTTTCAAGGATATGATAGTGGTGGTGC Rv CCGAATAATATATTCAATATATTCCTACTC	701
33F/33A/37	*wzy*	Fwd GAAGGCAATCAATGTGATTGTGTCGCG Rv CTTCAAAATGAAGATTATAGTACCCTTCTAC	338
34	*wzy*	Fwd GCTTTTGTAAGAGGAGATTATTTTCACCCAAC Rv CAATCCGACTAAGTCTTCAGTAAAAAACTTTAC	408
35A/35C/42	*wzx*	Fwd ATTACGACTCCTTATGTGACGCGCATA Rv CCAATCCCAAGATATATGCAACTAGGTT	280
35B	*werH*	Fwd GATAAGTCTGTTGTGGAGACTTAAAAAGAATG Rv CTTTCCAGATAATTACAGGTATTCCTGAAGCAAG	677
35F/47F	*wzy*	Fwd GAACATAGTCGCTATTGTATTTTATTTAAAGCAA Rv GACTAGGAGCATTATTCCTAGAGCGAGTAAACC	517
38/25F/25A	*wzy*	Fwd CGTTCTTTTATCTCACTGTATAGTATCTTTATG Rv ATGTTTGAATTAAAGCTAACGTAACAATCC	574
39	*wzy*	Fwd TCATTGTATTAACCCTATGCTTTATTGGTG Rv GAGTATCTCCATTGTATTGAAATCTACCAA	98
Capsular Polysaccharide A	*cpsA*	Fwd GCAGTACAGCAGTTTGTTGGACTGACC Rv GAATATTTTCATTATCAGTCCCAGTC	160

2.5.3 Differentiating serotypes

In silico methods of capsular typing can result in multiple serotype possibilities. In order to differentiate between possible serotypes, slide agglutination was performed to confirm the definitive serotype.

2.5.4 Differentiating serotypes 11A/D and 15B/C

Serotypes 11A/D, and 15B/C featured frequently in initial analyses. Availability of WGS data allowed for quick and robust identification of the definitive serotype by using ipcress (*In silico* PCR experiment Simulation System) to retrieve genes associated with serotype differentiation.

For identification of 11A and 11D, the *wcrL* gene was retrieved using ipcress and published primers (Oliver *et al.*, 2013a) and translated to an amino acid sequence

using an online translation tool from http://www.bioinformatics.org/sms2/translate.html, set to reading frame 1 and translation table 11. Clustal Omega (Sievers *et al.*, 2011) was then used to align the amino acid sequences with 11A and 11D sequences deposited in the National Centre for Biotechnology Information (NCBI) database (GenBank JX102570 and JX102571, respectively).

Differentiation of serotypes 15B and 15C was achieved by using ipcress to retrieve the O-acetlytransferase genes with primers published by van Selm *et alet al* (van Selm *et al.*, 2003b). The number of TA repeats within the gene indicates the serotype and functionality of the isolate. 15B has eight TA repeats and 15C has either seven or nine repeats which results in a loss of functionality.

Table 9 Primers used to differentiate 11A/D and 15B/C

Gene	Primer Sequence (5′ – 3′)	Reference
wcrL	Fwd ATGATACCTAAAAAGATTCATTATTG Rv TCATCTAGTTTTCCCCTTTAATA	Oliver *et al*, 2013
O-acetyl transferase	Fwd ATTTTGTTAAATAGGTAGGAAAG Rv TTCTTCTTTATCCGAACAGGC	van Selm *et al*, 2003b

2.5.1 Serotyping of year 10 and IPD isolates

The length of this study meant that developments in serotyping methods were able to be utilised when they became available. In 2015/16 I switched serotyping methods and used PneumoCaT (Pneumococcal Capsular Typing) (Kapatai *et al.*, 2016) for all of the year 10 and IPD isolates. PneumoCaT is an *in silico* tool that designates serotype by comparing capsular sequences with reference serotypes. Its advantage over similar tools is that it can distinguish molecular variations which can be applied to an algorithm to differentiate closely matched capsular types.

The change in *in silico* methods was validated by performing retrospective analyses on previously serotyped isolates. The emergence of genetic variants within serogroups prompted the re-typing of more than 100 pneumococci from

study years 1 to 9. PneumoCaT results were 100% concordant with previous methods.

2.5.2 MLST

Sequence typing was performed using SRST2 (Inouye *et al.*, 2014) which mapped the FASTQ files against an MLST database which held sequences of seven pneumococcal 'housekeeping' genes: *aroE; gdh; gki; recP; spi; xpt* and *ddl.* Sequence type (ST) designation is predefined and held on a curated public database (pubmlst.org/spneumoniae/).

2.5.3 Clonal complex

Sequence types were uploaded to eBURST (Feil *et al.*, 2004) an on-line tool that displays a set of STs in a graphical format. Clonal complexes (CC), defined as a set of STS with six out of seven identical alleles were identified and predictions were made regarding founding and sub-founding STs, i.e. those STs that displayed the most number of shared alleles with other STS. CCs were visualized using Phyloviz, http://www.phyloviz.net (Francisco *et al.*, 2012).

2.5.4 Genomic analysis

Production of core genome SNP trees were produced using parSNP, a tool of the Harvest suite (Treangen *et al.*, 2014). It aligns the core genomes of selected isolates, identifies single nucleotide polymorphisms (SNPs) and uses that to build a phylogenetic tree. An output of the Harvest suite is a file that can be used by a companion tool, Gingr, to interrogate the analysed genomes.

The core genome trees are in a newick format which can then be used as an input for visualisation tools. For this thesis I used Microreact (Argimón *et al.*, 2016) which uses the newick tree and corresponding data uploaded as a csv file to produce a range of visuals.

2.6 Strain exclusion

Five isolates failed the WGS quality checks carried out by the WTSI and were excluded from any further analysis. Two of these isolates: 9174 and 9227 displayed optochin resistance and altered morphology when re-cultured and were permanently removed from the dataset.

The assembly quality was checked for all isolates. Six carriage isolates had a high number of contigs and were therefore excluded from any analysis that required assembly data. The total genome length was too high in one carriage isolate form year 8 and one from year 9 and these too were removed from genomic analyses. All the remaining isolates had acceptable read lengths, number of contigs and N50.

2.7 Statistical analyses

This study uses descriptive data to assess frequencies of events. The number of frequencies has been used to determine percentages with all confidence intervals (CI) calculated at 95% confidence.

To compare one measure with another, the Fisher's exact test was used to produce a two-tailed p-value. To compare a series of measurements the chi square test for trend was used. All statistical tests set a p-value of <0.05 as the level of significance.

Odds ratios (OR) used to calculate the invasive potential of serotypes followed the methods laid out by Brueggemann *et alet al* (Brueggemann *et al.*, 2003) in that OR= (ad)/(bc) where a = number of IPD isolates of serotype X; b= number of carriage isolates of serotype X; c= number of IPD isolates not serotype X and d = number of carriage isolates not serotype X.

Serotype and sequence type diversity (D) was measured using the Simpson's Diversity Index (SDI) 1-D, using $D= \Sigma n(n-1)/N(N-1)$ where n= total number of a particular subset of a population and N= the total number of the population. Results fall between 0 and 1 and the higher the value the greater the diversity. To estimate if a range of values fell within expected levels of normal variation the D'Agostino and Pearson omnibus normality test was used.

All statistical analyses were undertaken using GraphPad Prism v7.03 (GraphPad Software, La Jolla California USA, www.graphpad.com).

IPD incidence was calculated using population data using Hampshire County Council census data 2011 (http://www3.hants.gov.uk/factsandfigures/population-statistics/census_pages/census_2011.htm). Southampton General Hospital serves the people of Southampton and South Hampshire therefore population data for Southampton and the New Forest area was estimated to be representative.

Census data are broken into age categories which formed the basis for the selection of ages in this study. Incidence rates were calculated per 100,000 to allow comparison with other studies and PHE data for England and Wales.

Calculations involving serotype specific IPD incidence had to be adjusted to allow for missing data. The method chosen was utilised by PHE and based upon assumptions that missing data matched known data. All adjustments were made by age group and by year to ensure results were as accurate as possible. Previous studies published by PHE have used this method when only 48% of data were available. As this study had available data for 71.5% of isolates it was deemed an acceptable method.

Chapter 3 *Streptococcus pneumoniae* isolated from carriage: 2006 – 2016

3.1 Introduction

Bacterial carriage studies provide information to ascertain the risk of disease and evaluate vaccine efficacy. This is especially important for the organism *Streptococcus pneumoniae* as not only is it responsible for a huge burden of morbidity and mortality, but also clinical interventions such as the implementation of pneumococcal conjugate vaccines (PCVs) have been shown to cause a rapid shift in circulating bacterial populations (Croucher *et al.*, 2011a).

Pneumococcal carriage studies conventionally focus on the colonisation of young children as this is where carriage has been observed to be greatest (Bogaert *et al.*, 2004a), however, while this has been well studied, the differences that exist within this important age group are often neglected. Identifying where pneumococcal carriage prevalence is greatest can contribute to our understanding of epidemiology and aid evaluation of vaccine impact.

Further, the implementation of PCV13 occurred in 2010 therefore studies conducted prior to 2015 that utilise data from the 0 to 4 years age group may have captured recipients of PCV13's predecessor PCV7. Stratification of this age group allows the impact of PCV13 to be measured with greater accuracy in participants of different ages or vaccine status.

Examination of pneumococcal carriage over a ten year period that encompasses the introduction of PCV13 provides an opportunity to deepen our knowledge of the impact of vaccine intervention.

3.1.1 Main aims

In order to understand how carriage contributes to disease, the population of isolates found in pneumococcal carriage must first be accurately characterised. The introduction of PCV13 must be examined due to the potential impact it is likely to have on the pneumococcal population. This is further complicated by the mixed age of carriage study participants as different ages will hold a different vaccine status. This chapter seeks to examine the impact of PCV13 on

pneumococcal carriage over a ten year period. The following points will be investigated.

- Changes in pneumococcal carriage rate over ten years
- Age stratification of pneumococcal carriers and comparison of age groups
- Vaccine status and pneumococcal carriage
- Serotype distribution of pneumococcal carriage over 10 years
- Examination of serotype distribution differences between age groups

3.2 Methods

The Southampton paediatric pneumococcal carriage study began in the winter of 2006 and has collected pneumococci carried by participants aged 0 to 4 years every winter since the study began. See Table 1 for my contribution to the study and methods 2.1.2 for the study design. Methods for study years 1 to 7 (winter 2006/07 to winter 2012/13) have been previously published (Tocheva *et al.*, 2011a, Gladstone *et al.*, 2015, Devine *et al.*, 2017).

Serotyping for years 8 to 10 has been carried out using four methods: three *in silico* methods, SRST2 and iPCRess (study years 8 and 9) and PneumoCaT (year 10) and one phenotypic method, latex agglutination. All isolates in years 8 and 9 underwent latex agglutination and in year 10 this method was employed only for ambiguous results. See methods 2.5 for description.

Age of participants was calculated for the first four years of the study using dates of birth given on the consent forms. For years five to ten participant ages were recorded from questionnaire data and relied on approximations from participant guardians.

3.3 Results: Carriage

3.3.1 Carriage study participants

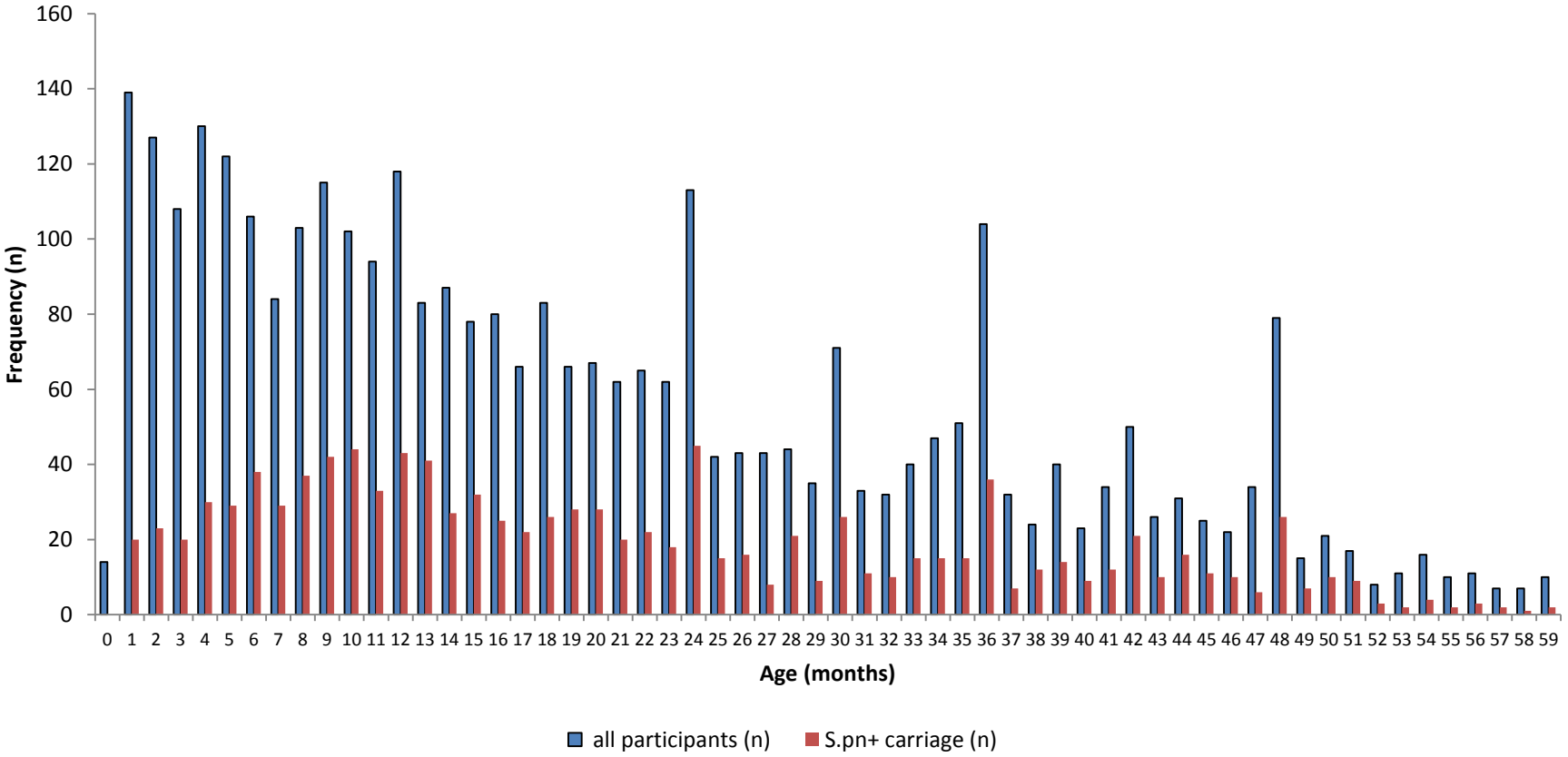
For the ten years from winter 2006/07 to winter 2015/16 a total of 3446 children aged four years and under participated in the carriage study. The age was recorded for 3412 of the children. The age of 34 participants was unknown and therefore excluded from analyses requiring age stratification. Figure 3 shows the age distribution of all study participants (blue bars).

The lowest number of study participants was in year 7 (winter 2012/13) when 218 children were recruited. The highest number of participants was recorded in year 10 (2015/16) where 521 children were recruited. An average number of 341 participants were recruited for each study year.

Microbiological testing of all isolates was conducted and the age distribution of participants from which pneumococci were isolated is given in Figure 3 (red bars).

The average age of study participants (from all 10 years) was 20.1 months (median= 17 months) and the average age of pneumococcal carriers was 21.5 months (median= 18 months).

Figure 3 Age distribution of study participants and pneumococcal carriers



Age of every study participant from winter 2006/07 to winter 2015/16 is represented by the blue bars. Age of every participant from which pneumococci were isolated is represented by the red bars.

3.3.2 Pneumococcal carriage

Pneumococcal carriage was calculated for each year of study and is represented in Figure 4.

In the first study year (winter 2006/07) 32% (26.9% to 37.1%) of participants were found to be carrying pneumococci and in year 10 pneumococci were isolated from 36.5% (32.4% to 40.6%) of participants. Chi square test for trend showed that carriage increased significantly over the ten years of the study (p= 0.001).

The lowest carriage rate was seen in study year 4 (winter 2009/10) when 27.8% (23.4% to 32.2%) of participants were found to be carrying pneumococci. The highest carriage rate was seen in year 9 (winter 2014/15) when 38.5% (33.4% to 43.6%) of participants carried pneumococci.

Figure 4 Pneumococcal carriage rate per study year

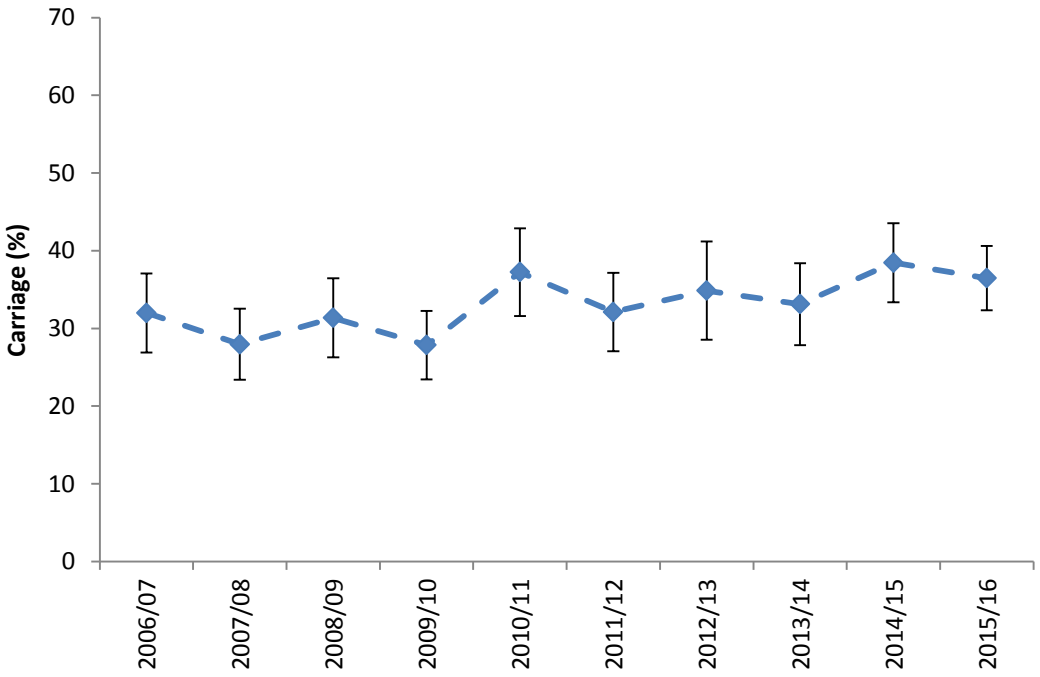


Figure 4 shows the carriage rate for each year of study. Pneumococcal carriage increased over the period of the study (X^2 test for trend p = 0.001).

3.3.3 Age stratification.

Study participants were separated into strata according to ages 0 to 11 months; 12 to 23 months; 24 to 35 months and 36 to 59 months. Those aged more than 36 months were merged into one stratum to compensate for the lower number of older participants and add power to statistical analyses. The numbers for each age strata are shown in Table 10.

Table 10 Number of participants in each age strata

	Total participants (n)			
Study year	0 to 11 mths	12 to 23 mths	24 to 35 mths	36 to 59 mths
2006/07	118	77	70	57
2007/08	89	112	66	105
2008/09	79	87	63	90
2009/10	194	114	57	30
2010/11	107	72	43	60
2011/12	119	87	48	73
2012/13	87	73	42	16
2013/14	103	81	62	59
2014/15	138	83	61	69
2015/16	210	131	82	98

3.3.4 Carriage rate by age

The pneumococcal carriage rate for each age strata was calculated and shown in Figure 5 (a-d).

Pneumococcal carriage in participants aged 0 to 11 months remained statistically stable over the 10 year study period (p = 0.9). As PCV13 was introduced in April 2010, participants aged 0 to 11 months would have received PCV7 in the first four years of the study (winter 2006/07 to 2009/10). The winter of 2010/11 is considered a transition year as some participants will have received PCV7 while others will have received PCV13 depending on month of birth. For study year 6 (winter 2011/12) onwards, all participants aged 0 to 11 months would've received PCV13 if they were up to date with their vaccination schedule. Table 11 details the study years that constitute pre and post PCV13 for this age group.

Pneumococcal carriage in participants aged 12 to 23 months increased significantly (p= 0.02) over the 10 year study period. The lowest carriage rate was in study year 6 (winter 2011/12) where 26.4% (17.1% - 35.7%) of participants were found to be carrying pneumococci and the highest carriage rate was in year 9 (winter 2014/15) where 47% (36.3% - 57.7%) of participants carried pneumococci. Participants of this age group would've been recipients of PCV7 if they received their vaccinations at the age recommended by the National Immunisation Programme (NIP) which in the UK is 8 weeks, 16 weeks and a booster at one year. Therefore post-PCV13 analysis of this age group would not be possible until after study year 7 (winter 2012/13). Study year 6 was considered a transition year where participants may have received either PCV7 or PCV13 depending on when they were born. Table 11 details the pre and post PCV13 study years for this age group.

For participants aged 24 to 35 months pneumococcal carriage increased significantly over the 10 study years (p= 0.0005). Carriage was lowest in study year 2 (winter 2007/08) with 16.7% (7.7% - 25.7%) of participants found to be carrying pneumococci. Carriage was highest in study year 9 (winter 2014/15) with 45.9% (33.4% - 58.4%) of participants carrying pneumococci. Pre PCV13 study years for this age group is defined as study years 1 to 6 (winter 2006/07 to 2011/12) as all participants of this age group would've been recipients of PCV7 if their vaccinations were given at the recommended time-points. Study year 7 was considered a transition year as participants may have received either PCV7 or

PCV13 depending on their month of birth. Post PCV13 analysis could be calculated from year 8 (winter 2013/14) onwards.

Participants of the age group 35 to 59 months were found to have a statistically stable rate of pneumococcal carriage over the study period (p= 0.5). The pre-PCV13 period could be defined as study years 1 to 7 (winters 2006/07 to winter 2012/13) as participants of this age group would've received PCV7 if vaccination was given at the recommended time-points. Due to the merging of older participants into one age strata, the transition period between pre and post PCV13 eras encompassed two study years, 8 and 9. Only study year 10 (winter 2015/16) could be used to analyse post PCV13 in this age group. Table 11 details the pre and post PCV13 study periods.

Figure 5 (a–d) Pneumococcal carriage per age strata

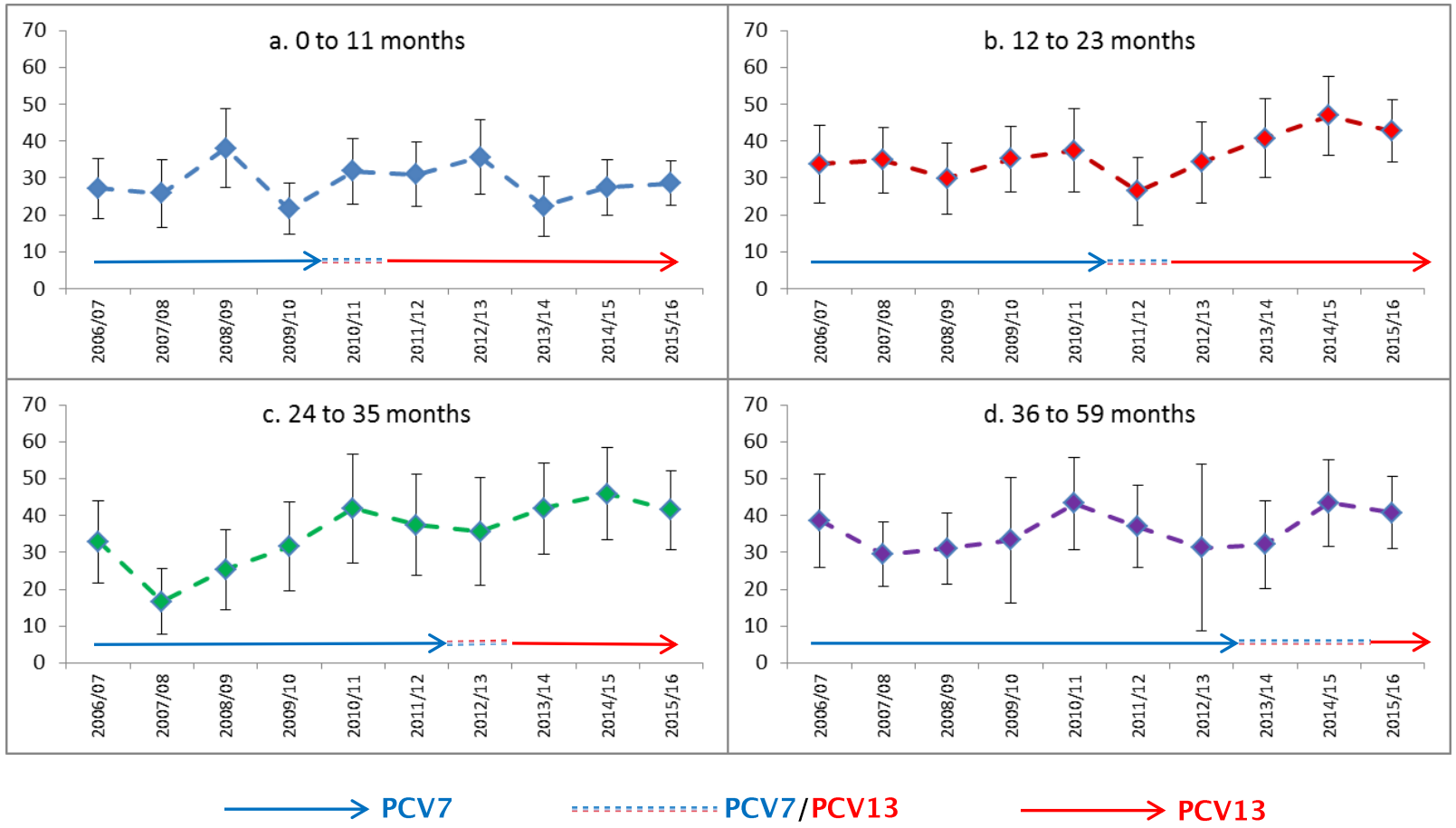


Figure 5 a–d show the pneumococcal carriage rate per age group. PCV7 or PCV13 administration is dependent on participant age and indicated by coloured arrows.

Table 11 pre and post PCV13 study years by age group

Age Group	Pre PCV13	Post PCV13
0 to 11 months	Years 1 to 4 (2006/07 to 2009/10)	Years 6 to 10 (2011/12 to 2015/16)
12 to 23 months	Years 1 to 5 (2006/07 to 2010/11)	Years 7 to 10 (2012/13 to 2015/16)
24 to 35 months	Years 1 to 6 (2006/07 to 2011/12)	Years 8 to 10 2013/14 to 2015/16
36 to 59 months	Years 1 to 7 (2006/07 to 2012/13)	Year 10 2015/16

To investigate the impact of PCV13 in participants aged 0 to 59 months, different study years would need to be investigated dependent on participant age. Study years 1 to 4 (2006/07 to 2009/10) are the only years that capture the pre PCV13 pneumococcal population in all participants. Study year 10 (2015/16) is the only year which can be used for post PCV13 analysis if all participants aged 0 to 59 months are used.

3.3.5 PCV13 impact on pneumococcal carriage

In order to examine the impact of PCV13 on pneumococcal carriage the pre and post PCV13 carriage rate was calculated for each age group.

A comparison of different methods for measurement is made in Figure 6 (a-c). The first method uses pre and post PCV13 eras as defined by age group. For participants aged 0 to 11 months, pre PCV13 is represented by study years 1 to 4, year 5 is excluded and post PCV13 uses years 6 to 10. For the 12 to 23 month age group, study years 1 to 5 are included, year 6 is excluded and years 7 to 10 is considered post PCV13. For participants aged 24 to 35 months, years 1 to 6 represents the pre-PCV13 era, year 7 is excluded and years 8 to10 represents post PCV13. Finally for the 36 to 59 month age group, Years 1 to 7 can be included as pre PCV13, years 8 and 9 are excluded and year 10 is used to measure post PCV13 carriage.

Using this method, pneumococcal carriage increased in all age groups from pre PCV13 to post PCV13. The largest change was seen in the 24 to 35 month age group which increased from a pre PCV13 carriage rate of 30% to a post PCV13 carriage rate of 42.9% (p= 0.002). The 12 to 23 month age group also saw a significant increase from 34.2% pre PCV13 to 41.8% post PCV13 (p= 0.03). Whilst an increase in carriage was seen for both age groups, 0 to 11 months and 36 to 59 months, this was statistically non-significant with p-values of 0.4 and 0.3, respectively.

The analysis was repeated however the same study years were used for each of the age groups. In study years 1 to 4 all of the participants, regardless of age would have been recipients of PCV7 and therefore these years were used to measure pre PCV13 pneumococcal carriage. For post PCV13, only year 10 could be included in the analysis as this is the only year where all study participants would have received PCV13.

Using the same study years to compare carriage in all age groups, only the 24 to 35 month group showed a significant increase in carriage from a pre PCV13 rate of 26.6% and a post PCV13 rate of 41.5% (p= 0.01). An increase in carriage was seen in all other age groups however they did not appear to be statistically significant. For 0 to 11 months, p= 0.6, 12 to 23 months p= 0.07 and for 36 to 59 months p= 0.2).

Finally, PCV13 impact was examined again using study years 1 to 4 for all age groups and study years 7 to 10 for post PCV13 in all age groups. Whilst this method captures recipients of PCV7 in the post PCV13 group it serves as a comparison to other pneumococcal studies that use similar years to examine the impact of PCV13.

This method revealed smaller increases in pneumococcal carriage from the pre PCV13 era to the post PCV13 era. Only the 24 to 35 month group showed a statistically significant increase from 26.6% to 41% (p= 0.0004). The 12 to 23 month group had a non-significant increase from 33.6% to 38.7% (p= 0.1) as did the 0 to 11 month group (p= 0.4) and the 36 to 59 month group (p= 0.1).

Figure 6 (a–c) pre and post PCV13 carriage by age group using three methods for measurement

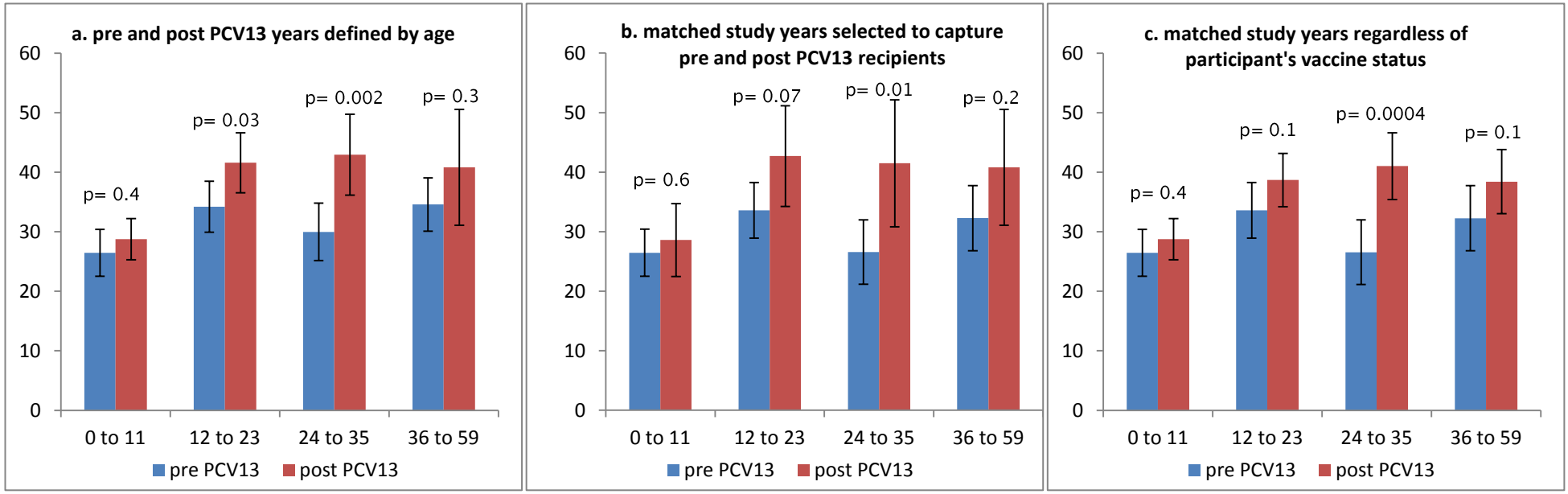


Figure 6 (a–c) shows the results of analyses using three different methods for measuring the impact of PCV13 on pneumococcal carriage. In Figure 6a pre PCV13 is defined as the study years during which participants of each age group would have received PCV7 and post PCV13 uses only those study years where participants would have received PCV13. Figure 6b uses matched study years (pre PCV13 = study years 1 to 4 and post PCV13= study year 10) and Figure 6c also uses matched study years however it does not take into account whether older participants were recipients of PCV7 or PCV13. An increase in post PCV13 pneumococcal carriage was seen in all age groups however in Figure 6a two age groups; 12 to 23 months and 24 to 35 months show a statistically significant increase following PCV13 introduction. Only the 24 to 35 month age group showed a statistically significant increase in carriage in every analysis.

3.3.6 Vaccine status

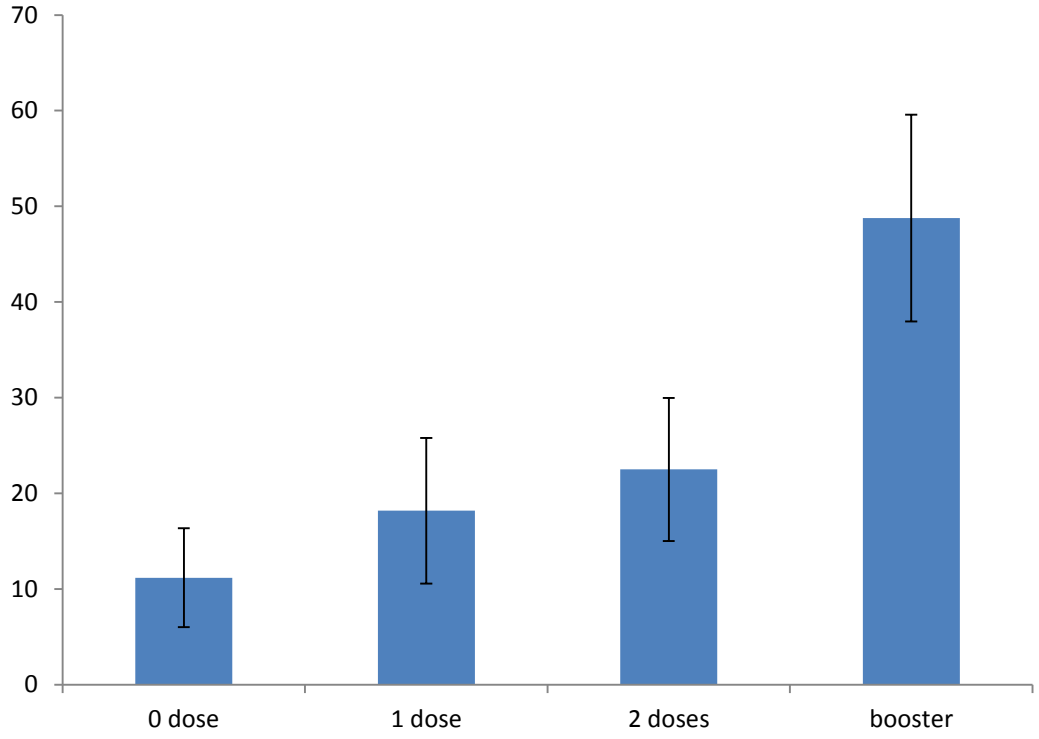
The vaccine status of study participants was elicited from age and questionnaire data. From study year 5 (2010/11) questionnaires that were completed for every participant provided information on vaccine status. For study years 1 to 4, vaccine status was inferred from participant age.

Study participants aged less than 2 months were categorized as having no dose of PCV. From study year 5 this status was confirmed using questionnaire data. To analyse pneumococcal carriage after one dose PCV, participants aged 3 months who were recorded as being up to date with their vaccinations were used and for 2 doses participants aged 5 months and up to date with routine vaccinations were used. For the booster category participants aged 13 months and up to date with their vaccinations were used in the analysis. Table 12 shows the number of participants in each group category.

Table 12 Vaccine status of study participants

Vaccine status	Total number (all years)	Pneumococcal carriage (all years)
No dose (<2 months)	143	16
1 dose (3 months)	99	18
2 doses (5 months)	120	27
Booster (13 months)	82	40

Figure 7 Carriage rate by PCV vaccine status

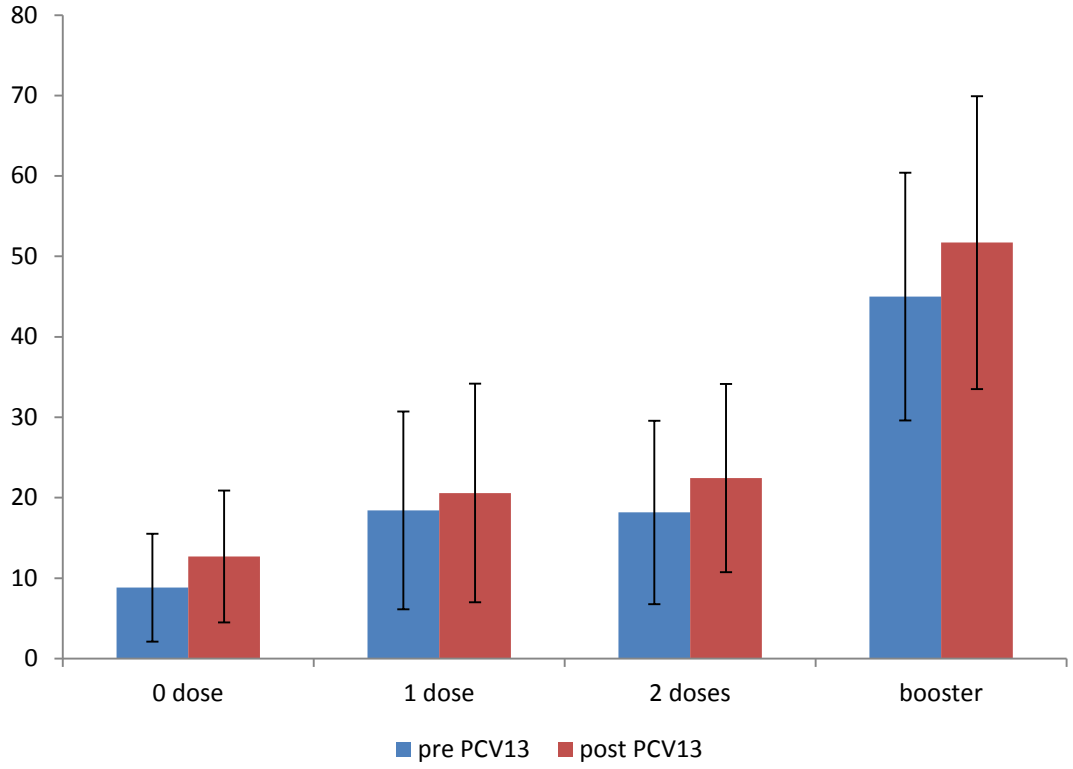


The lowest carriage rate was found in participants that hadn't received any doses of PCV (11.2% ±5.2). Carriage increased to 18% ± 7.6 (p= 0.1) in participants that had received one dose of PCV and then increased to 22% ± 7.6 (p= 0.5) in participants that had received 2 doses. The carriage rate increased significantly (p= 0.0001) to its highest rate of 48.8% ± 10.8 in participants that had received the PCV booster.

To assess the impact of PCV13 on groups of different vaccine status, the pre and post PCV13 carriage rate was calculated for each category and is shown in Figure 8.

A post PCV13 increase in carriage was observed in all categories. In participants that had never received a dose of PCV, carriage increased from 8.8% to 12.7% (p= 0.6), for participants that had received 1 dose PCV there was a small increase in carriage from 18.4% to 20.6% (p= 1). Participants who had received two doses had a pre PCV13 carriage rate of 18.2% and a post PCV13 rate of 22.4% (p= 0.8) and finally carriage in participants who had received the PCV booster had an increase in carriage from 45% to 51.7% (p= 0.6).

Figure 8 pre and post PCV13 carriage per vaccine status



3.3.7 Live Attenuated Influenza Vaccine Status and Pneumococcal Carriage

Questionnaire data allowed examination of the effects that other vaccines might have on pneumococcal carriage. In 2013, the UK began a phased introduction of a live, attenuated influenza vaccine (LAIV) called Fluenz Tetra® which is administered to children aged between two and four years of age in an effort to lessen the burden of seasonal flu. LAIV is a nasal spray containing live, weakened forms of the influenza virus. Once administered, LAIV mimics a natural viral infection by replicating within the nasopharynx promoting an immunological response in the recipient (Hoft *et al.*, 2017).

The study questionnaire was updated in 2014 to identify participants that had received this vaccine. To reduce the confounding effect of differing carriage rates by age, only participants aged between two and four years were used in the analysis. Of the 872 study participants that provided questionnaires in the winters of 2014/15 and 2015/16 (years 9 and 10 of study) 310 of them were aged between 2 and 4 years. Of these 134 had received Fluenz Tetra®, 157 had not and 19 were unsure of their status. Pneumococcal carriage was higher (47%) in those that had received LAIV than those that had not (36.9%), (p= 0.09).

Figure 9 LAIV status and pneumococcal carriage

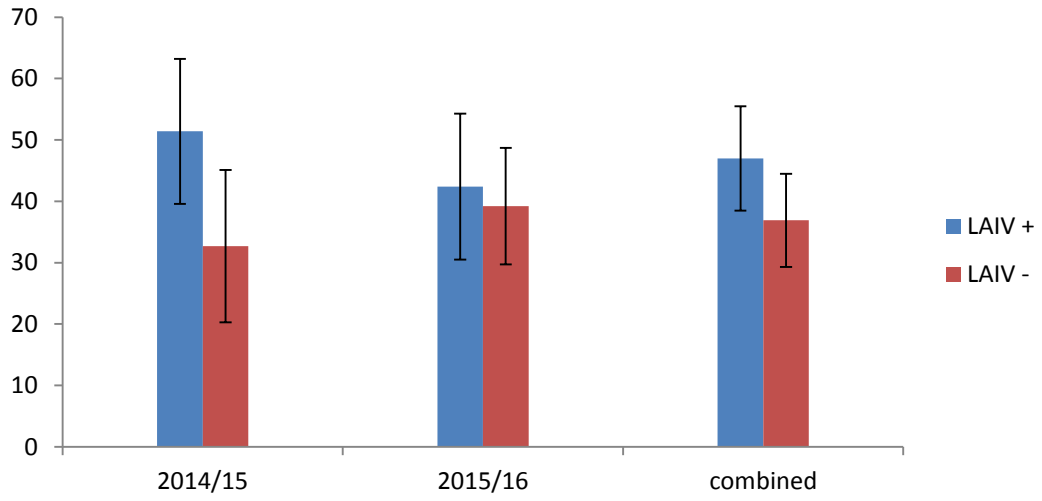


Figure 9 shows that for each of the winters (2014/15 and 2015/16) pneumococcal carriage was higher in LAIV recipients than in non-recipients.

3.3.8 Serotype distribution

For each year of the study approximately 100 pneumococcal isolates were serotyped. Table 13 shows the exact number that were analysed each year.

Table 13 Pneumococcal isolates available for serotype analysis

Year	Total S.pn isolates	Exclusions	Number serotyped	Known serotype and participant age (n)
2006/07	104	1 excluded by QC.	103	102
2007/08	104		104	104
2008/09	102		102	100
2009/10	111		111	110
2010/11	102	4 failed to re-grow	98	97
2011/12	105	7 failed extraction/ sequencing	98	98
2012/13	78		76	76
2013/14	101		101	101
2014/15	135	35 to be sequenced	100	100
2015/16	189	89 to be sequenced	100	100
Totals	**1131**	**136 (124/136 to be sequenced)**	**993**	**988**

3.3.9 Serotype distribution by study year

A total of forty-five different serotypes were identified from 993 pneumococcal isolates. Eight of these serotypes constituted >50% of the pneumococcal population: 15B/C (n= 99), 11A (n= 89), 23B (n= 66), 6C (n= 61), 21 (n=55), 35F (n= 49), 22F (n= 49) and 15A (n= 47). Eight serotypes were observed in every year of the study which included seven of the above; 15B/C, 11A, 23B, 35F, 21, 22F and 6C plus serotype 23A. Due to frequent phase variation between serotypes 15B and 15C they are grouped together underneath the designation 15B/C, serotype 21A and 21B are grouped within serogroup 21 and genotype 23B1 is grouped within serotype 23B.

The impact of PCV13 on serotype distribution was examined by comparing the prevalence of individual serotypes in the study years before PCV13 introduction with serotype prevalence post PCV13. As participants of all ages were included, the pre PCV13 era was represented by study years 1 to 4 (winter 2006/07 to winter 2009/10) and study year 10 (winter 2015/16) was used to capture post PCV13 serotype distribution. Study years 5 to 9 (winter 2010/11 to winter 2014/15) were excluded to ensure all participants had received the same valency vaccine.

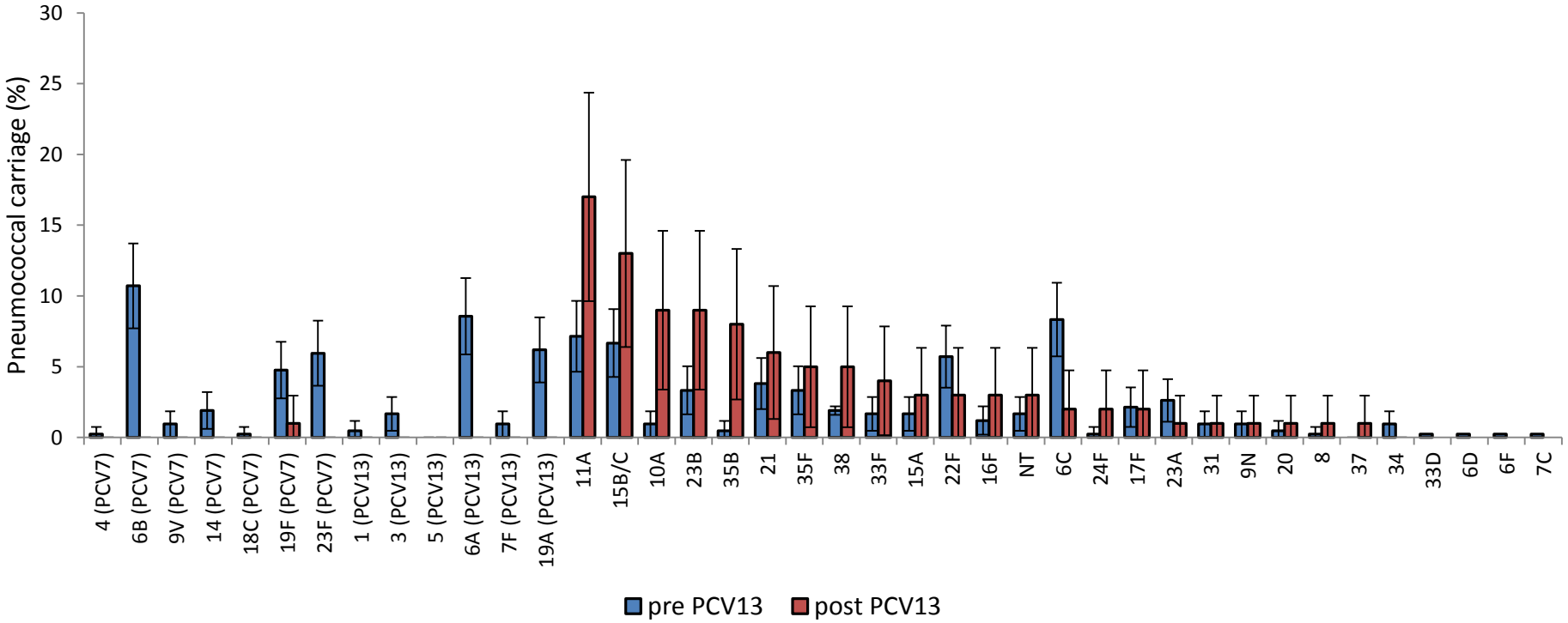
Examination of individual serotype frequency over time revealed that ten serotypes had a statistically significant change in prevalence. Five serotypes: 6B, 23F (included in PCV7/13), 6A, 19A (included in PCV13) and 6C were observed to decrease while five serotypes: 11A, 15B/C, 10A, 23B and 35B (all non-vaccine serotypes), showed an increase and are presented in Table 14. All serotypes found in study years 1 to 4 are compared to the serotypes observed in study year 10 and presented in Figure 10.

Table 14 Serotypes with significant changes in prevalence

Serotype	PCV7/PCV13 NVT	Change	Pre PCV13[2] (%)	95% CI	Post PCV13[3] (%)	95% CI	p-value[4]
6B	PCV7/13	↓	10.7	± 3	0	-	<0.0001
23F	PCV7/13	↓	6	± 2.3	0	-	0.007
6A	PCV13	↓	8.6	± 2.7	0	-	0.0006
19A	PCV13	↓	6.2	± 2.3	0	-	0.005
11A	NVT	↑	7.1	± 2.5	17	± 7.4	0.006
15B/C	NVT	↑	6.7	± 2.4	13	± 6.6	0.04
10A	NVT	↑	1	± 0.9	9	± 5.6	<0.0001
23B	NVT	↑	3.3	± 1.7	9	± 5.6	0.03
35B	NVT	↑	3.3	± 1.7	5	± 4.3	<0.0001
6C	NVT	↓	8.3	± 2.6	2	± 2.7	0.03

[2] Pre PCV13 data derived from study years 1 to 4 (2006/07 to 2009/10)
[3] Post PCV13 derived from study year 10 (2015/16) to ensure capture of PCV13 recipients only (not PCV7)
[4] P-values calculated using Fisher's exact test

Figure 10 Serotype prevalence pre PCV13[5] and post PCV13[6]

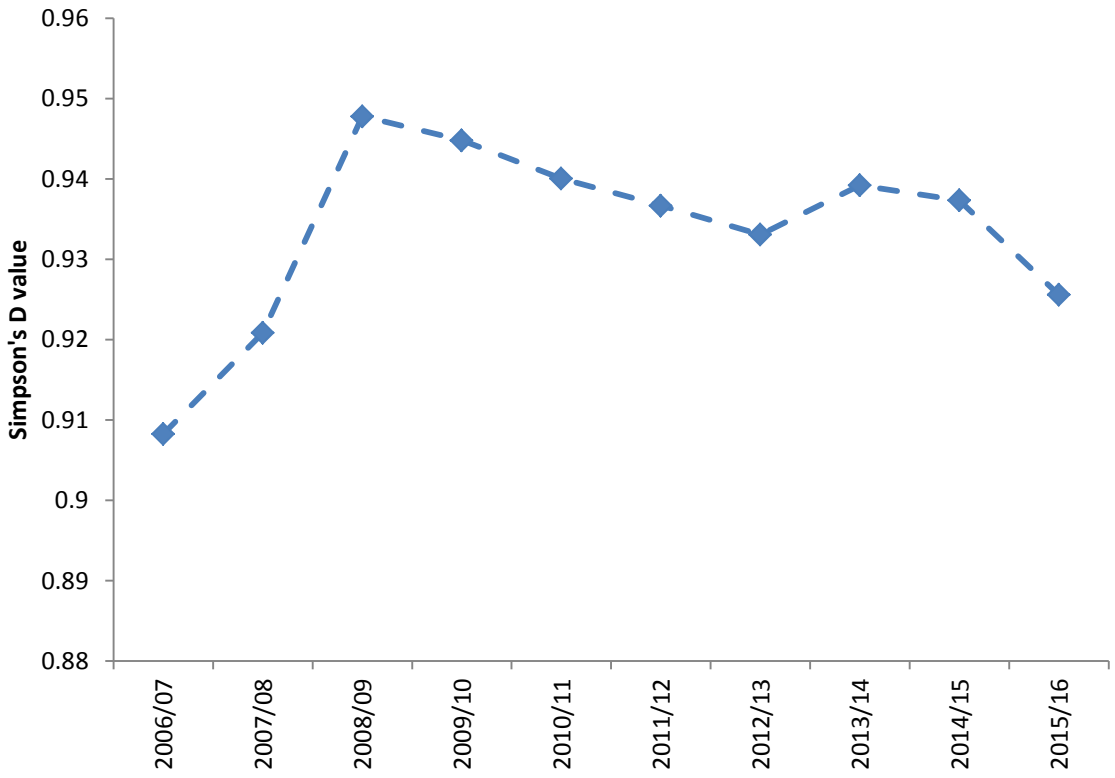


[5] Pre PCV13 data derived from study years 1 to 4 (2006/07 to 2009/10) and are represented by the blue bars.
[6] Post PCV13 data derived from study year 10 (2015/16) and are represented by the red bars.
Error bars represent 95% confidence interval.

3.3.10 Serotype diversity

In order to assess the degree of serotype diversity within the pneumococcal population, the Simpson diversity index (1-D) was used to quantify diversity according to the number of serotypes present and their relative abundance. The Simpson Index gives a value (D) between 0 and 1, which in this case represents the probability that two randomly selected isolates will belong to the same serotype. A D value of 0.95 was determined for the ten year study period indicating that this population has a high degree of diversity. D value calculations for each study year showed the least amount of diversity in study year 1 (0.91) and the most amount of diversity in study year 3 (0.95). The D values for each study year is shown in Figure 11.

Figure 11 Serotype diversity per study year



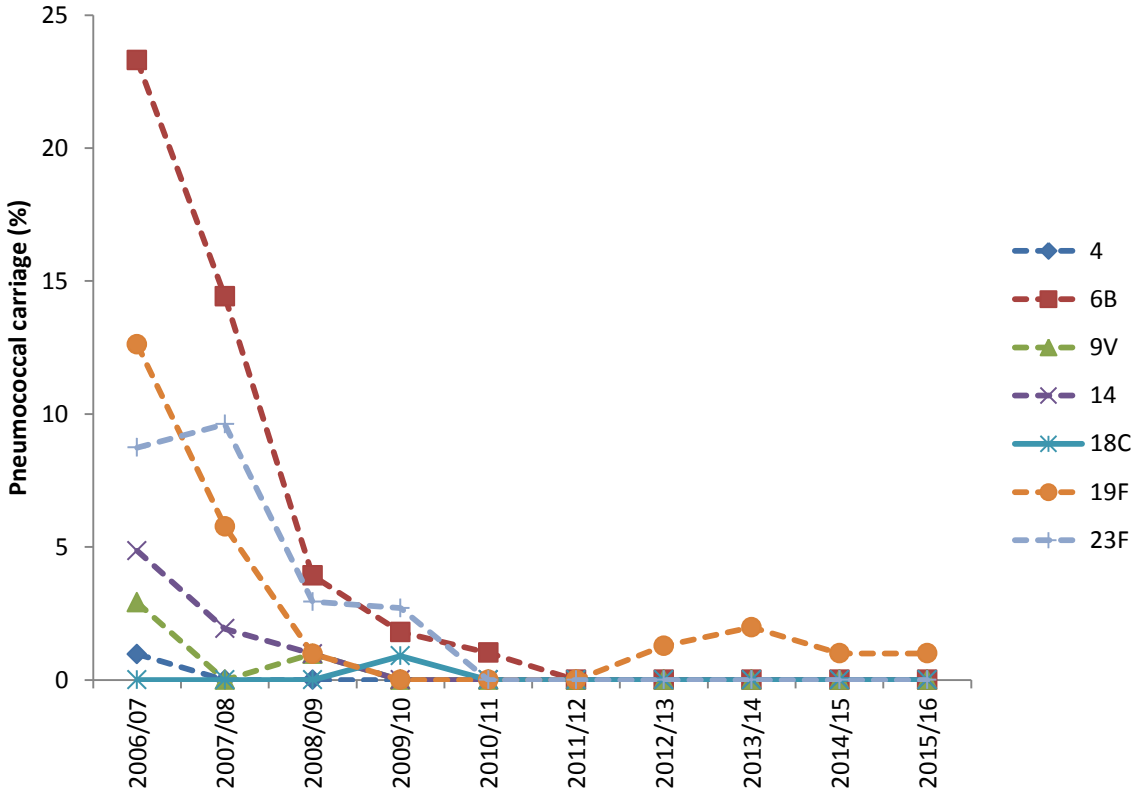
3.3.11 PCV7 serotypes in carriage

Carriage of PCV7 vaccine-type (VT) serotypes (4, 6B, 9V, 14, 18C, 19F and 23F) in study year 1 were: 4 (1%), 6B (23%), 9V (2.9%), 14 (4.9%), 18C (0%), 19F (11.8%), and 23F (8.8%). In total this constituted 52% (95% CI 42.3% – 61.7%) of all pneumococcal carriage.

By study year 10, carriage of PCV7 serotypes had significantly decreased (p=<0.0001) to 1% (0.97% – 2.97%) of the pneumococcal population with one isolate of 19F being recorded at this time-point. **Error! Reference source not ound.** shows the carriage of PCV7 VT serotypes over the ten year study period.

Serotype 19F was the only PCV7 VT observed in carriage after 2010/11. It was isolated in 2012/13 (n= 1), 2013/14 (n= 2), 2014/15 (n= 1) and 2015/16 (n=1).

Figure 12 Carriage of PCV7 serotypes in children aged four years and under: 2006/07 to 2015/16



3.3.12 PCV13 serotypes in carriage

In 2010, the PCV7 vaccine was updated to include an additional six serotypes: 1, 3, 5, 6A, 7F, and 19A. The combined carriage rate of these six serotypes was 16.5% (9.3% – 23.7%) in the winter of 2006/07 and rose to 21.4% (13.5% – 29.3%) in the fourth year of the study (p= 0.5) which was the winter that preceded the introduction of PCV13.

Following the implementation of PCV13, a decrease in carriage rate was observed for the six additional serotypes. By year 10 (2015/16), none of the additional six serotypes in PCV13 were observed in pneumococcal carriage. The carriage of PCV13 serotypes over 10 years is shown in Figure 13.

The decrease in carriage was measured from both the first study year (p= 0.002) and from study year 4 (p= <0.0001) which was the winter prior to PCV13 introduction. The decrease in PCV13 serotypes was accompanied by a concomitant increase in non-PCV13 serotypes which is shown in Figure 14.

Three PCV13 VTs were observed in years eight and nine of the study: serotype 3 (n=1), 19A (n=3) and 7F (n=1). An increase in 19A carriage was observed (p=0.07) before the introduction of PCV13 which then decreased to a lower rate of carriage in subsequent years. In the ninth year of study, 19A represented 2% (95% CI -0.8% – 4.8%) of the pneumococcal population which was a statistically significant decrease (p=0.01) from its peak in 2009/10.

Figure 13 Carriage of PCV13 serotypes in children aged four years and under: 2006/07 to 2015/16

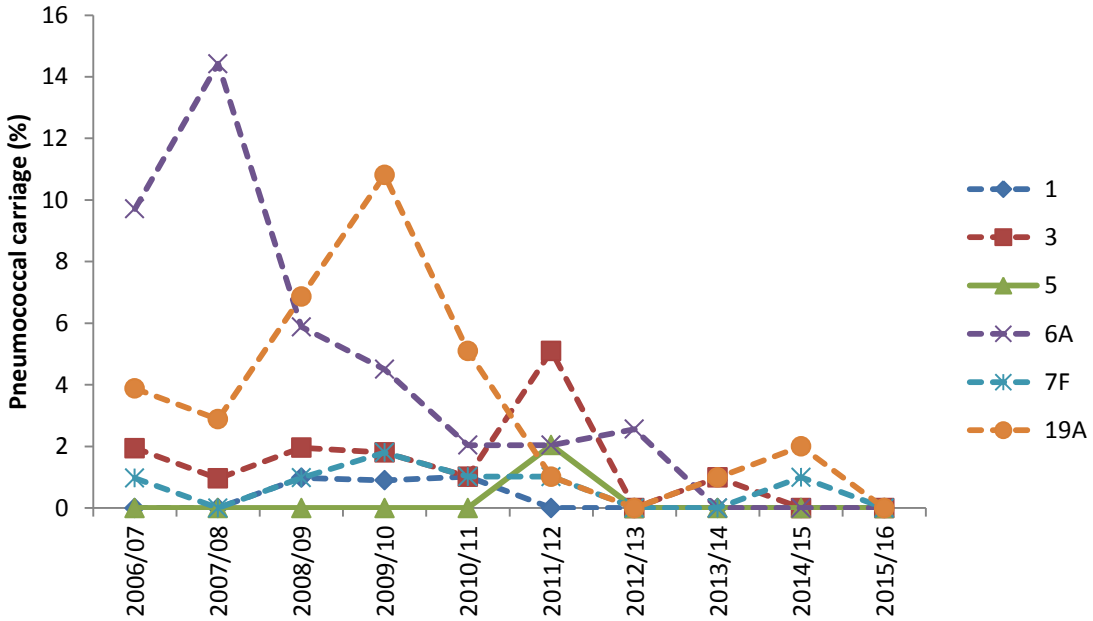


Figure 14 Carriage of vaccine type (VT) and non-vaccine type (NVT) serotypes in children aged four years and under: 2006/07 to 2015/16

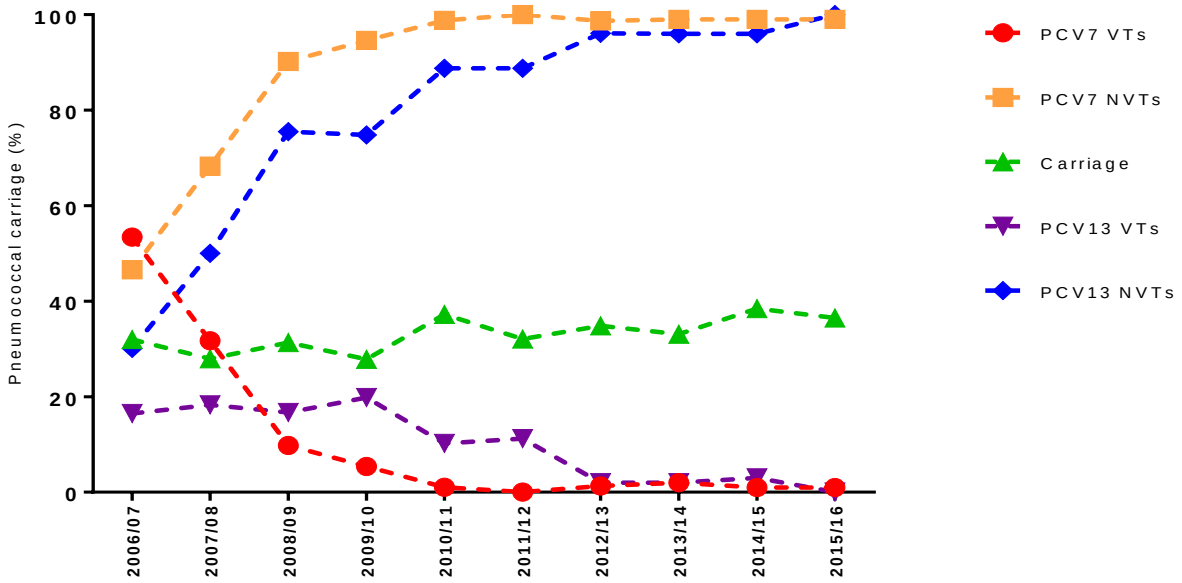


Figure 14 shows the decrease of PCV7 and PCV13[7] VT serotypes over the study period. A concomitant increase in NVT serotypes was observed over the same time period. PCV7 NVTs constitute all serotypes not included in PCV7. PCV13 NVTs constitute all serotypes not included in PCV13.

[7] PCV13 VT serotypes are the additional six serotypes included in PCV13 but not in PCV7 (1, 3, 5, 6A, 7F and 19A).

3.3.1 Non-vaccine serotypes in carriage

The decrease in VT serotypes was accompanied by a concomitant increase in non-PCV7 serotypes which is shown in Figure 14. The carriage rate for non-PCV7 serotypes in year 1 was calculated at 48% (38.3% – 57.5%) and this increased to 99% (93% – 100%) in year 10 (p=<0.0001).

The largest increase in carriage was seen in NVT serotypes 11A, 15B/C, 10A, 23B and 35B. Carriage of these serotypes equated to 9.7% (4% – 15.6%) in year 1 and rose to 56% (46.3% – 65.7%) of carriage in year 10 (p=<0.0001). Figure 15 shows the carriage of these serotypes over 10 years.

Temporal fluctuations of individual serotypes were observed over the study period. An increase in serotype 6C was previously reported (Tocheva *et al.*, 2010, Loman *et al.*, 2013) however 6C decreased from 13.7% of carriage in year 3 to 3.6% of carriage in year 4 (p= 0.01) before increasing again in study year 4 (9.2%). 6C then decreased and was observed in 2% of the carriage isolates in study year 10 (Figure 16)

An increase in serotype 15A was also previously reported (Devine *et al.*, 2017) after it was observed to significantly increase from 1% of pneumococcal carriage in study year 1 to 12.8% of carriage in year 6. Since then, carriage of 15A has decreased and was observed to make up 3% of pneumococcal carriage in study year 10 (Figure 16).

Carriage of other individual serotypes was observed to fluctuate over the study period. Serotypes 21, 22F and 33F have had periods of increase and decrease over the 10 year study period and this observation is shown in Figure 17.

Figure 15 NVT serotypes that have increased significantly in carriage

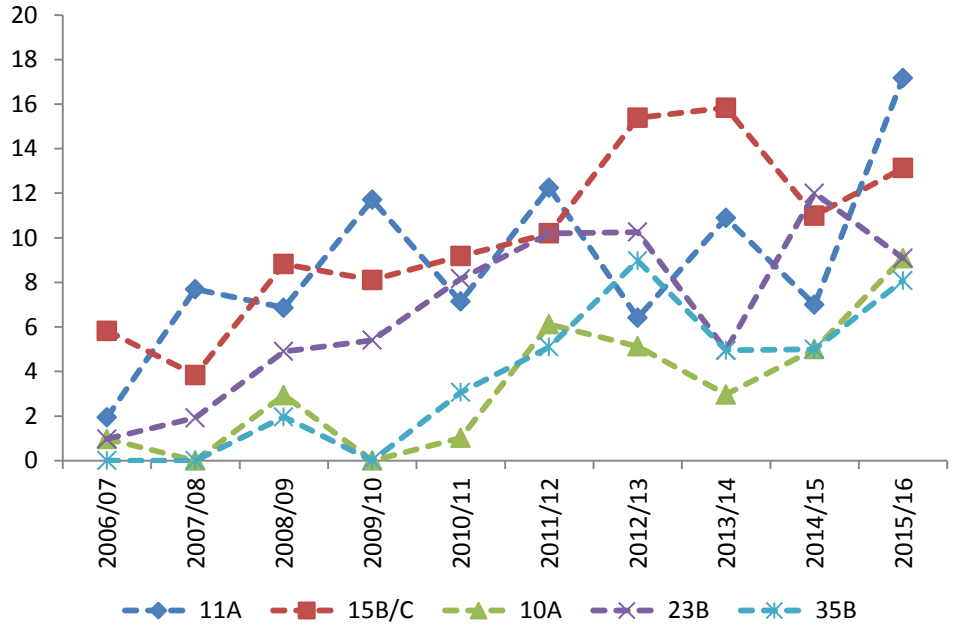
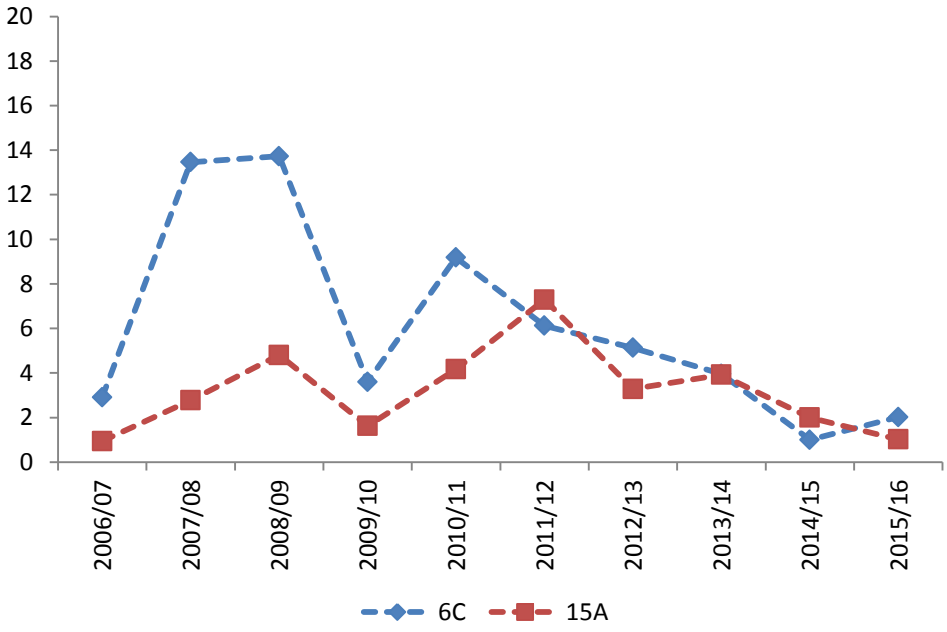


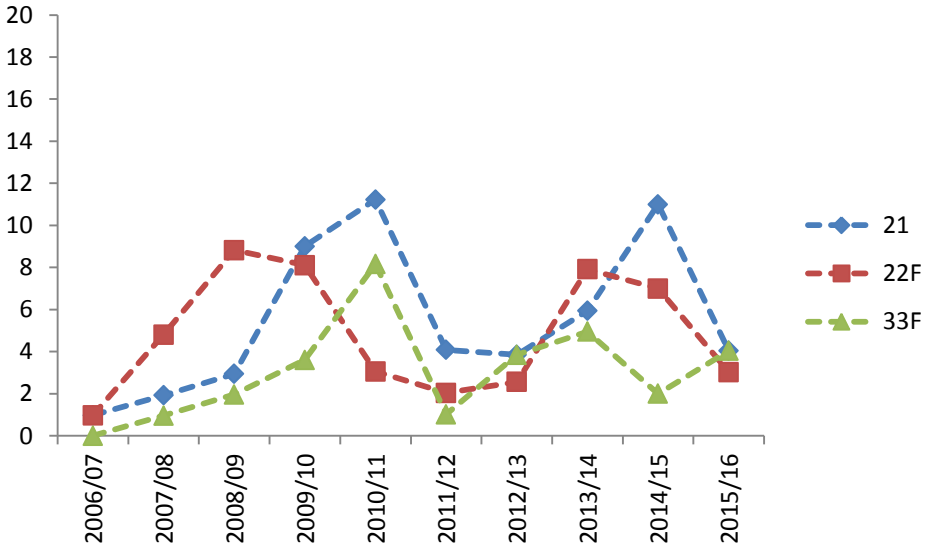
Figure 15 shows the serotypes that have increased significantly in carriage over the study period. They include 11A (p= 0.006), 15B/C (p= 0.04), 10A (p= <0.0001), 23B (p= 0.03) and 35B (p= <0.0001).

Figure 16 Temporal fluctuations of previously reported serotypes 6C and 15A



Increases in carriage of 6C and 15A have previously been reported however further observations over time show that the increase was not maintained.

Figure 17 Temporal fluctuations of NVT serotypes 21, 22F and 33F



Whilst small fluctuations were observed for many NVT serotypes, 21, 22F and 33F showed a number of significant increases and decreases over the 10 year study period.

3.3.2 Age and serotype prevalence

Prevalence of individual serotypes was analysed for each age group, 0 to 11 months, 12 to 23 months, 24 to 35 months and 36 to 59 months.

In participants aged 0 to 11 months, a significant decrease was seen in VT serotypes: 6B (p= <0.0001), 19F (p= 0.0001), 23F (p= 0.003), and 6A (p= 0.003). Whilst 6B was the VT serotype observed most overall (n= 18), serotype 19A was the VT serotype observed in the most study years (7/10 years).

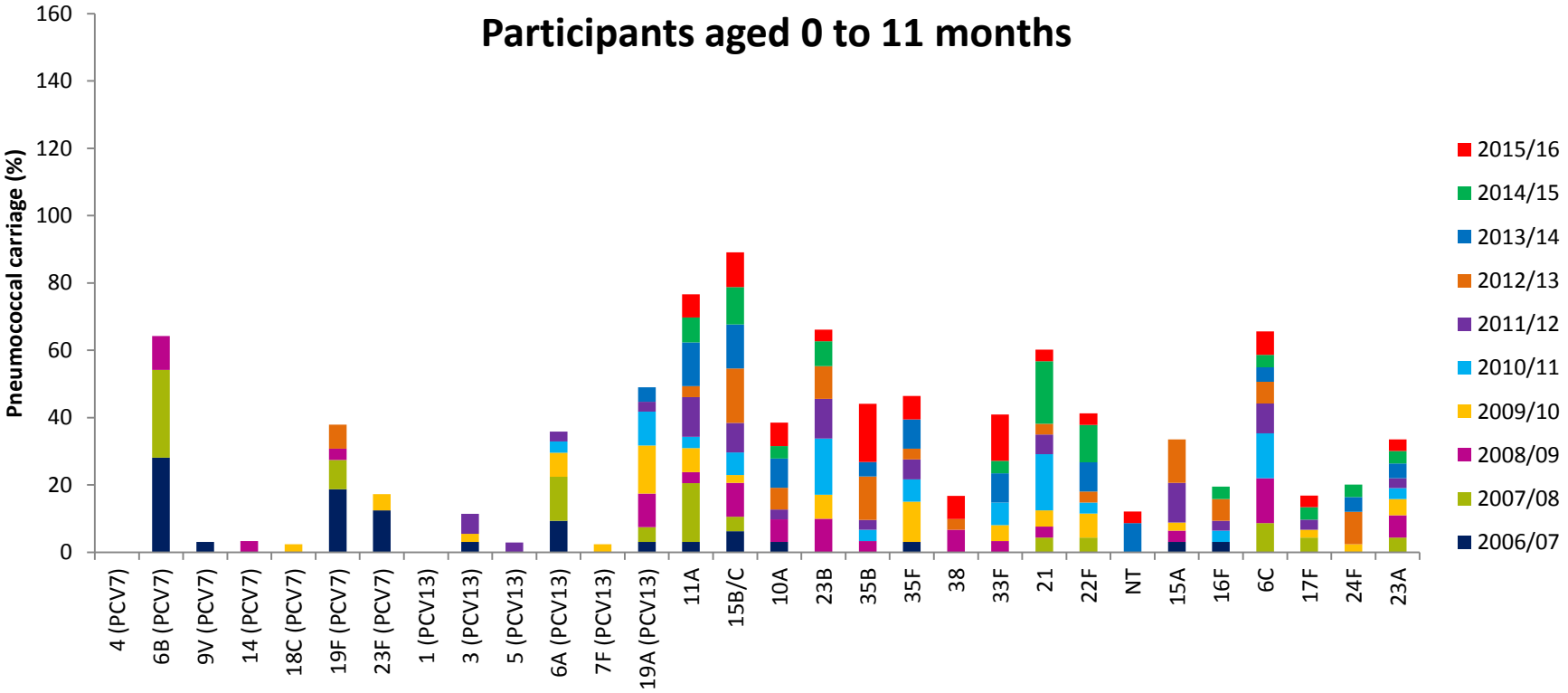
Two NVT serotypes were observed to increase significantly, 35B (p= 0.003) and 33F (p= 0.003). Prevalence of non-typeable (NT) pneumococci was also observed to increase (p= 0.04) in this age group. Serotypes 35B and 33F were the most frequently isolated serotypes in the 0 to 11 month olds in study year 10 (2015/16). Serotype distribution in participants aged 0 to 11 months is shown in Figure 18.

Observations of serotype prevalence in participants aged 12 to 23 months showed that six VT serotypes decreased significantly over the ten years: 14 (p= 0.03), 23F (p= 0.0007), 19A (p= 0.005), 6B (p= 0.0002), 19F (p= 0.008) and 6A (p= 0.002). A decrease was also observed in NVT 6C (p= 0.01) and 17F (p= 0.002). Five NVT serotypes were seen to increase significantly over ten years: 10A (p= 0.004), 23B (0.0007), 35B (p= 0.02), 16F (0.02) and 24F (p= 0.02) (Figure 19).

Over ten years, participants aged 24 to 35 months displayed a significant decrease in VT serotypes 6B (p= 0.0002), 14 (p= 0.002), 23F (p= 0.002) and 6A (p= 0.002) and NVT 6C also decreased (p= 0.007). Increases were observed in NVT serotypes 23B (p= 0.04) and 15A (p= 0.05) (Figure 20).

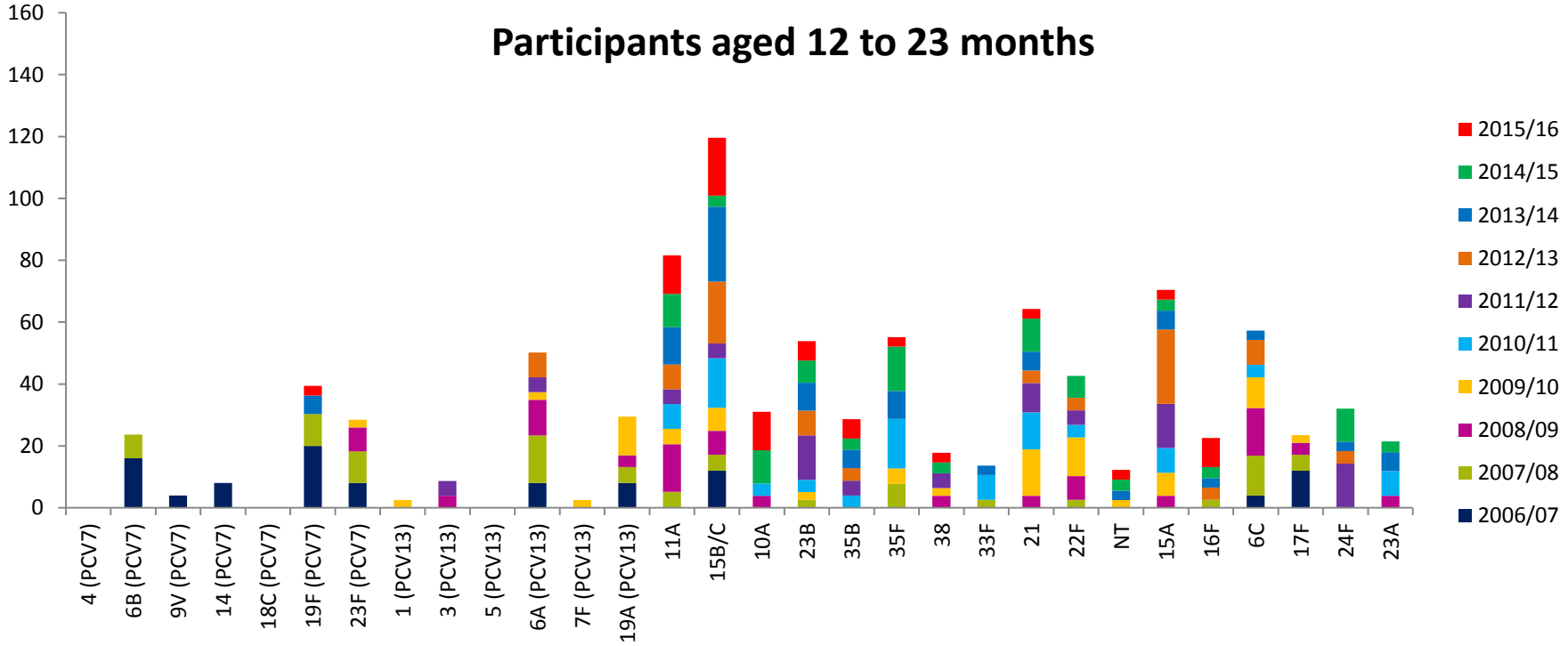
In participants aged 36 to 59 months, a significant decrease was seen in VT serotypes 6B (p= <0.0001), 14 (p= 0.04), 23F (p= 0.002) and 6A (p= 0.001). Significant increases were observed in NVT serotypes 10A (p= 0.04), 24F (p= 0.006) and 15A (p= 0.004) (Figure 21).

Figure 18 Serotype prevalence over ten years in participants aged 0 to 11 months



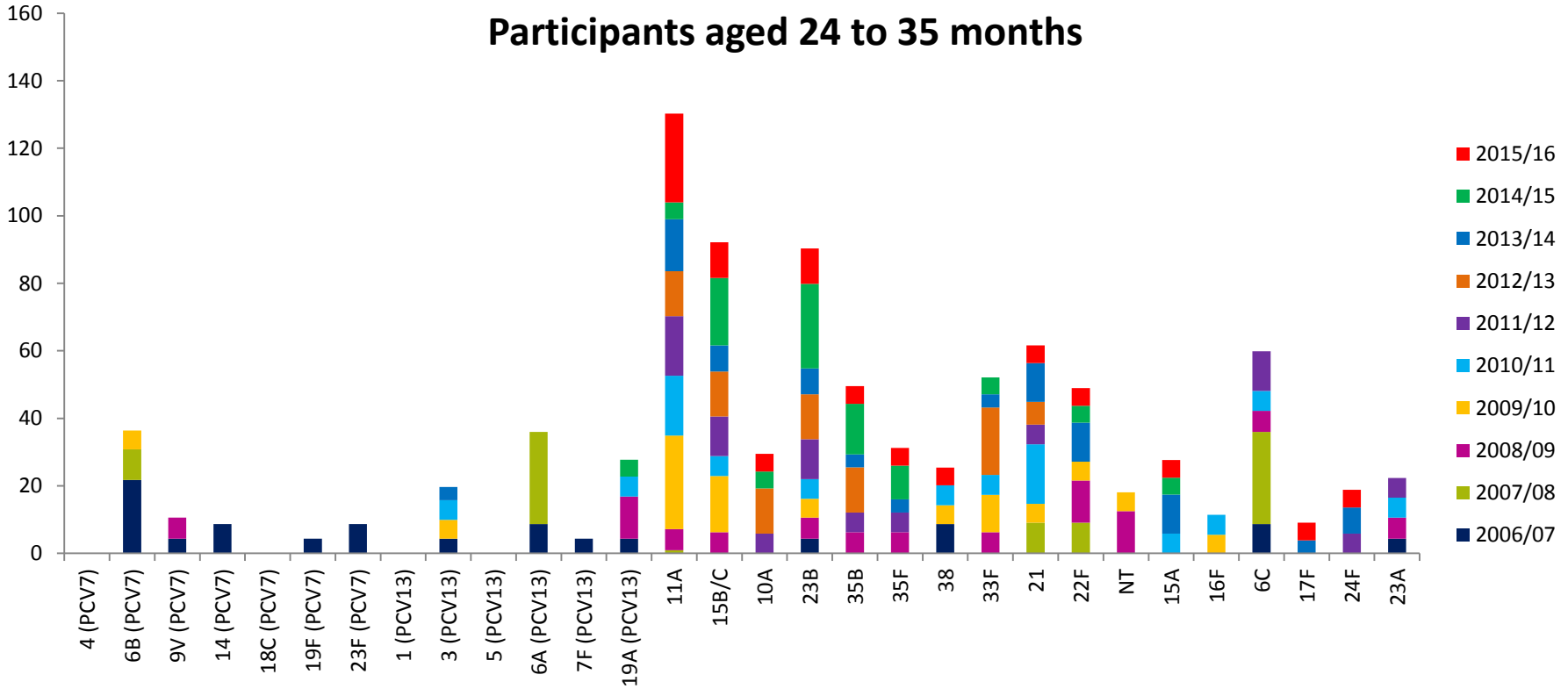
In participants aged 0 to 11 months, carriage was dominated by VT serotypes in study year 1 (78%). Serotype 19A was the VT that persisted the longest and was last isolated in study year 8 (2013/14). The most frequently isolated serotypes in year 10 were 35B and 33F, both of which have shown a significant increase in carriage by 0 to 11 month olds over the 10 year study period. Overall, serotypes 11A and 15B/C have been observed most frequently and they have been isolated in every year of the study.

Figure 19 Serotype prevalence over ten years in participants aged 12 to 23 months



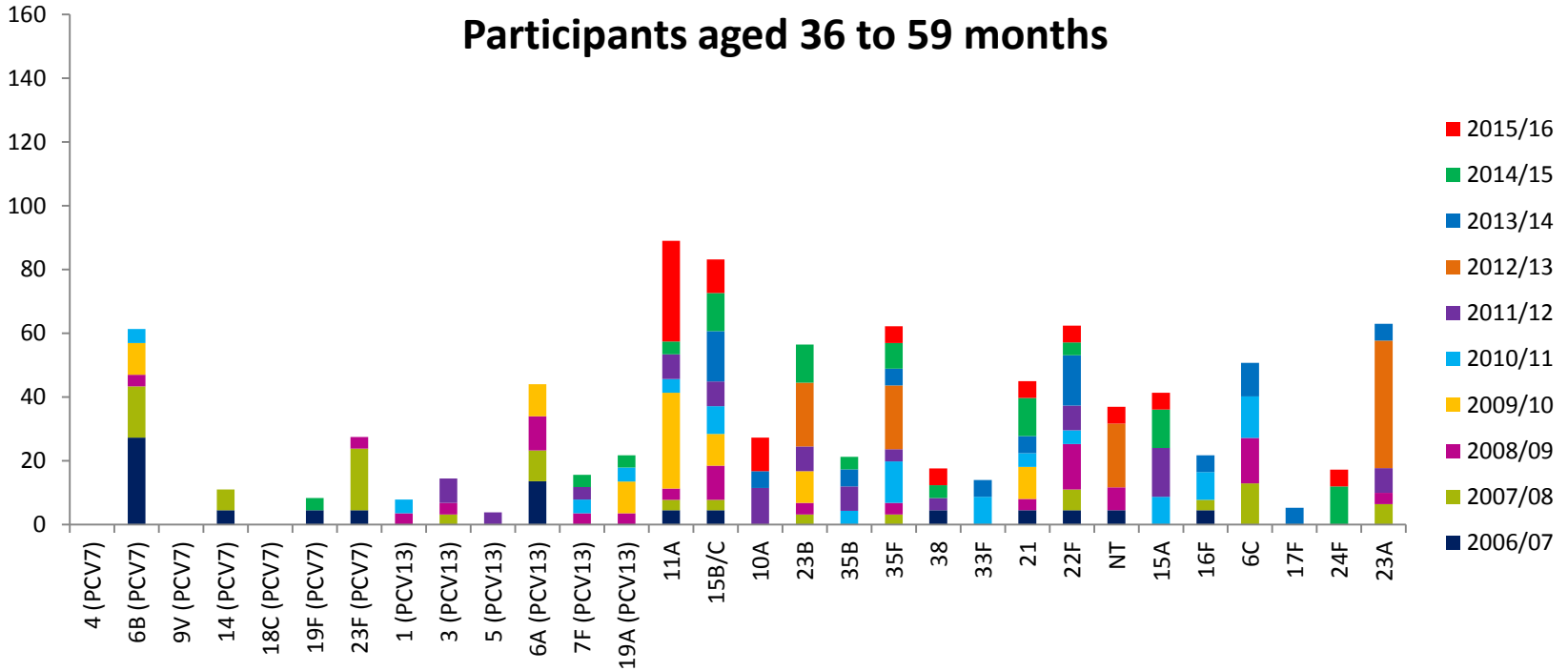
In participants aged 12 to 23 months carriage was dominated by VT serotypes in study year 1 (72%). Serotype 19F was the most recent VT serotype found in carriage amongst this age group (isolated in 2015/16). The most frequently isolated serotypes in year 10 were 11A, 15B/C and 10A. Overall, serotypes 11A and 15B/C have been observed most frequently and 15B/C has been isolated in every year of the study.

Figure 20 Serotype prevalence over ten years in participants aged 24 to 35 months



In participants aged 24 to 35 months carriage was dominated by VT serotypes in study year 1 (69.6%). Serotype 19A was the most recent VT serotype found in carriage amongst this age group (isolated in 2014/15). The most frequently isolated serotypes in year 10 were 11A, 15B/C and 23B. Serotypes 11A and 15B/C have been observed most frequently overall although they were not observed in this age group until year 3 (2008/09).

Figure 21 Serotype prevalence over ten years in participants aged 36 to 59 months



In participants aged 36 to 59 months carriage of VT serotypes in study year 1 was observed in just over half of isolates (54%). Serotypes 19F and 7F were the most recent VT serotypes found in carriage amongst this age group (isolated in 2014/15). The most frequently isolated serotypes in year 10 were 11A, 15B/C and 10A and these made up over 50% of all isolates.

3.4 Discussion

It was widely believed that the introduction of the pneumococcal conjugate vaccination into the national immunisation programme would lower the overall carriage rate via the removal of prevalent serotypes. Surveillance studies, published over the last decade have reported no such decrease in the carriage rate and have attributed serotype replacement as the reason for continued pneumococcal pervasiveness (Miller *et al.*, 2011, Hanage *et al.*, 2010).

The carriage study reported here shows that not only did the pneumococcal carriage rate not decrease following the routine use of PCVs but increased significantly ($p= 0.001$) over the 10 year study period.

There are a number of factors that need to be considered in order to understand this result. The introduction of PCVs led to a decrease in VT serotypes which were replaced by NVT serotypes. Eight NVT serotypes were found in every year of the carriage study: 11A, 15B/C, 23B, 35F, 21, 22F, 6C and 23A. Overall these serotypes constituted 50% of all pneumococcal carriage. These serotypes were already circulating in the population when PCV13 was introduced and were therefore already established as successful colonisers. Their presence in the pre PCV13 population ensured they were in a good position to fill the niche following the removal of VT serotypes. However, if these serotypes were only replacements for VT serotypes then the overall carriage rate would be expected to remain stable, not increase as we have observed.

Other studies have shown that the pneumococcal capsular type is a major determinant for colonisation duration (Sleeman *et al.*, 2006) and colonisation density (Rodrigues *et al.*, 2016). Examination of serotype-specific colonisation density has shown that serotype 35B, which increased significantly over the 10 year study period, has a greater carriage density relative to other serotypes. However these studies give no indication that serotypes such as 11A and 15B/C have increased colonisation periods or are carried at greater density relative to other serotypes. Further studies, focussed on serotype-specific carriage duration and density, are necessary to fully understand whether the significant increase of serotypes 11A, 15B/C, 10A, 23B and 35B are due to colonisation characteristics or because of increased acquisition. It is also important to continue surveillance on these serotypes to examine if this trend continues and to assess the impact of this on disease.

An update to the study questionnaire in 2014 provided an opportunity to examine how vaccines other than PCV may impact pneumococcal carriage. The introduction of the live attenuated influenza vaccine (LAIV) during the course of the study prompted an analysis of its impact on pneumococcal carriage. The LAIV is administered pernasally and induces an immune response stimulated by the live virus replicating within the nasopharynx. Preliminary findings have found an increase in pneumococcal carriage amongst recipients of the LAIV. Data collected in 2016/17 (not included in this thesis) has shown that pneumococcal carriage in LAIV recipients has continued to be higher than in non-recipients. Overall the three years for which data are available, pneumococcal carriage is significantly higher ($p= 0.03$) than in non-recipients. Other studies have shown an increase in pneumococcal density as well as prolonged pneumococcal carriage following administration of LAIV (Mina *et al.*, 2014, Mina *et al.*, 2015, Thors *et al.*, 2016b). As the swabbing period of the carriage study coincides with the time-point at which LAIV is administered, it is possible that this is contributing to the increased carriage rate although it is undetermined whether this is due to changes in acquisition rates or prolonged carriage.

Age stratification of study participants has shown that the increase in pneumococcal carriage was not statistically significant in all age groups. Whilst carriage remained at a stable rate in the 0 to 11 month and 36 to 59 month age groups, carriage in the 12 to 23 month and the 24 to 35 month groups was observed to increase. Serotypes 11A, 15B/C, 10A and 23B which were observed to significantly increase in carriage overall were the most prevalent serotypes in these age groups. However this leaves us with a 'chicken and egg' scenario, where it is difficult to understand if the increase in carriage amongst these two age groups led to the increase in these serotypes or whether these serotypes caused an increase in specific age groups.

Understanding the differences in carriage rates between individual age groups is important to assess vaccine impact. Studies that use pre and post PCV13 time-points that do not take into consideration the true vaccine status of individual participants risk capturing mixed populations of vaccine recipients. Comparison of pre and post PCV13 carriage found that a significant increase in carriage was seen in two age groups (12 to 23 months and 24 to 35 months) when the pre and post PCV13 periods were defined by the participants' age. Repeated analysis using the PCV13 introduction date to divide the pre and post PCV13 periods altered the outcome and a significant increase was seen in only one age category (24 to 35 months).

Age of participant is also important to record when evaluating vaccine escape serotypes. In many studies, post PCV13 analyses are dated from the year following PCV13 introduction. If the analyses include participants aged 0 to 4 years then there is a risk of recording PCV13 vaccine escape from non-PCV13 recipients. In this study 21 VT serotypes were observed in all participants between study year 6 (one year after PCV13 introduction) and year 10. Of these, 16/21 represented serotypes covered by PCV13. Analysis of participant age showed that 9/16 participants were from older age categories and therefore unlikely to have received PCV13.

The extent to which study design and choice of analytical methods has affected findings is unknown. Changes that have been implemented over the 10 study years have the possibility of affecting the carriage rate. For the first six years of the study, participant swabs were processed in a diagnostic laboratory by technicians outside of the research group. From the seventh year, swabs were processed within the laboratory of the research group and while both diagnostic and research technicians followed the same procedure, it must be acknowledged that this change could have affected observations. Discussions surrounding future vaccine development and/or changes to the vaccination schedule are reliant on accurate data derived from carriage studies. In order to provide this, age and vaccine status of participants must be scrutinised in order to capture as true a representation of pneumococcal carriage as possible.

Chapter 4 *Streptococcus pneumoniae* isolated from cases of Invasive pneumococcal disease: 2006 – 2016

4.1 Introduction

The extent to which pneumococcal carriage in young children influences disease in the wider population is the focus for a number of epidemiology studies. Studies have shown a decrease in invasive pneumococcal disease (IPD) incidence rates following PCV7 introduction (Miller *et al.*, 2011) however the degree to which PCV13 has affected IPD remains a subject of interest.

Whilst national surveillance data relating to IPD is available, examination of IPD from a smaller area in which carriage data are also available can provide valuable insights into the pneumococcal transmission.

Understanding pneumococcal epidemiology relies on identifying where pneumococcal carriage is most prevalent, where the burden of disease lies and how these two populations relate to one another. To continue this process the IPD incidence rate will be examined and the serotype distribution of both carriage and IPD populations will be compared.

4.1.1 Main aims

In order to investigate the influence of pneumococcal carriage on IPD, the incidence of IPD over the same 10 years for which the carriage isolates were collected must be examined. As IPD data are not limited to one age group, it will also be useful to compare disease in younger age groups who are likely to be PCV recipients with older age groups.

This chapter seeks to examine IPD incidence in Southampton from 2006 to 2016 and characterise the pneumococcal isolates responsible for cases of IPD. Findings may then be compared to the pneumococcal carriage population that was characterised in the previous chapter. To this end, the following points will be investigated:

- IPD incidence rate over 10 years (2006 to 2016)
- the burden of disease in different age groups
- serotype-specific IPD cases over 10 years
- a comparison of carriage and IPD
- serotype-specific disease potential

4.2 Methods

IPD isolates were collected from the Public Health England (PHE) laboratory at Southampton General Hospital. The laboratory runs a diagnostic service for patients of the hospital and therefore identification and storage of the isolates was carried out by PHE staff. Stored isolates and patient metadata relating to the isolates have been made available to our research group from 2006 to present.

Stored IPD isolates that were originally isolated between July 2006 and June 2016 were included in this analysis. The stored isolates were re-grown for DNA extraction and sent to the Wellcome Trust Sanger Institute (WTSI) for sequencing. For serotyping, PneumoCaT was used in the first instance and latex agglutination employed for ambiguous results (see methods 2.5). Not all IPD isolates were saved or were able to be regrown. Table 15 outlines the number collected by PHE and the number that have been successfully sequenced. Duplicate isolates were defined as belonging to the same patient within 30 days. In total 33 isolates were identified as duplicates and removed from the analysis.

IPD isolates with missing serotype results were corrected using the methods employed by PHE in their national surveillance publications (Waight *et al.*, 2015, Ladhani *et al.*, 2018) whereby serotype distribution is assumed to be the same as those with complete data.. Each epidemiological year was analysed separately and patient data was used to ascertain age of host. Known serotypes for the same age groups were matched to cases with unknown serotypes, for example, if eight patients aged over 65 years were found to have serotypes 10A (n = 6) and 22F (n= 2), and one patient had an unknown serotype, the number for that year would be serotype 10A (n= 6.75) and 22F (n= 2.25).

IPD incidence rates were estimated using Office of National Statistics 2011 census data using population data for the Southampton and South Hampshire (New Forest) that covers the area served by University Hospital Southampton NHS Foundation trust (methods 2.7).

Table 15 IPD isolates collected 2006 – 2016

Year	Isolates collected by PHE (n)	Isolates with WGS data available (n)
2006/07	56	34
2007/08	43	37
2008/09	51	45
2009/10	44	38
2010/11	35	31
2011/12	32	7
2012/13	40	30
2013/14	24	6
2014/15	29	24
2015/16	50	34
Total	404	286

The total isolates collected by PHE in 10 years does not include the 33 isolates identified as being duplicates. Serotype data for 118 isolates was unavailable either due to non-storage of original isolate, failure to re-grow or unsuccessful QC checks.

4.3 Results: IPD

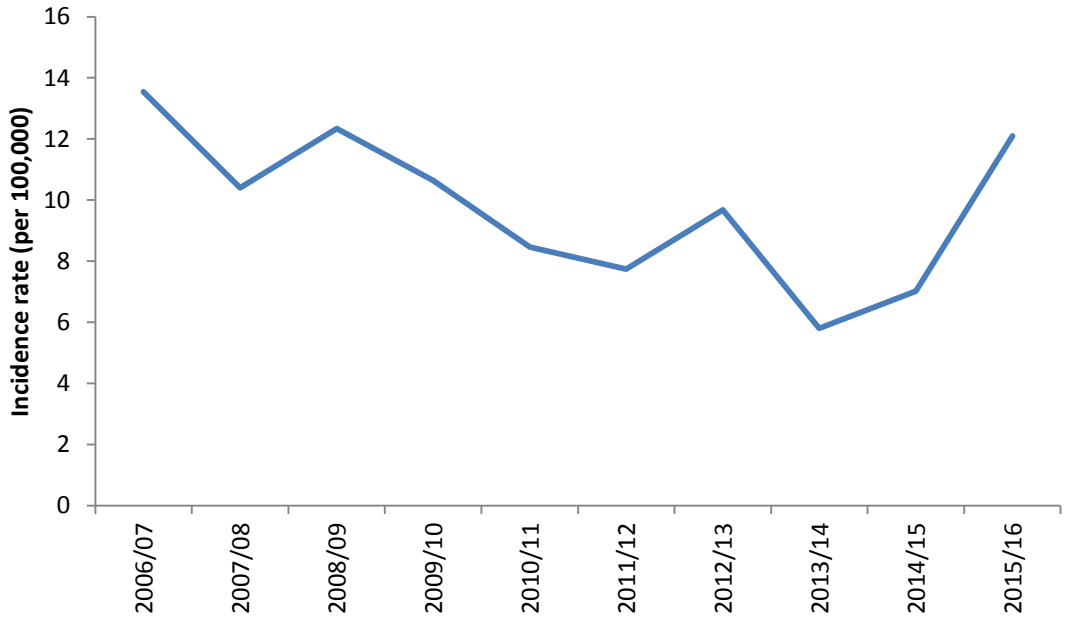
4.3.1 IPD incidence rate

IPD isolates were separated into 10 epidemiological years each running from July to June. The first year comprised of isolates collected between July 2006 and June 2007 and the last were collected between July 2015 and June 2016. A total of 404 isolates were collected by PHE during that time.

The incidence rate per year was calculated using population data from the Office of National Statistics 2011 census. A total population of 413344 was used to calculate IPD incidence rates. Figure 22 represents the IPD incidence rate for the Southampton and South Hants area over 10 years. Incidence rates were calculated per 100,000 to allow comparison with other studies and PHE data for England and Wales.

Overall, incidence of IPD was observed to decrease over the study period (p= 0.008) however a year by year analysis shows that IPD began to increase from its lowest incidence rate of 5.8 cases per 100,000 in the epidemiological year 2013/14 to 12.1 cases per 100,000 in 2015/16 (p= 0.003).

Figure 22 IPD incidence rates 2006 – 2016



4.3.2 Age distribution of IPD patients

Age data were available in 401/404 (99%) of IPD cases that were recorded between 2006 and 2016. The youngest patients had laboratory confirmed cases of IPD within 1 day of birth and the oldest was aged 99 years. The age distribution of all IPD patients is shown in Figure 23. A total of 286 IPD isolates were available for whole genome sequencing. The age distribution of these isolates is also included in Figure 23.

Patients were grouped into the age categories, 0 to 4 years, 5 to 14 years, 15 to 44 years, 45 to 64 years and 65+ years. Age groups were chosen to match the constructs of the 2011 census data to allow more robust incidence estimates.

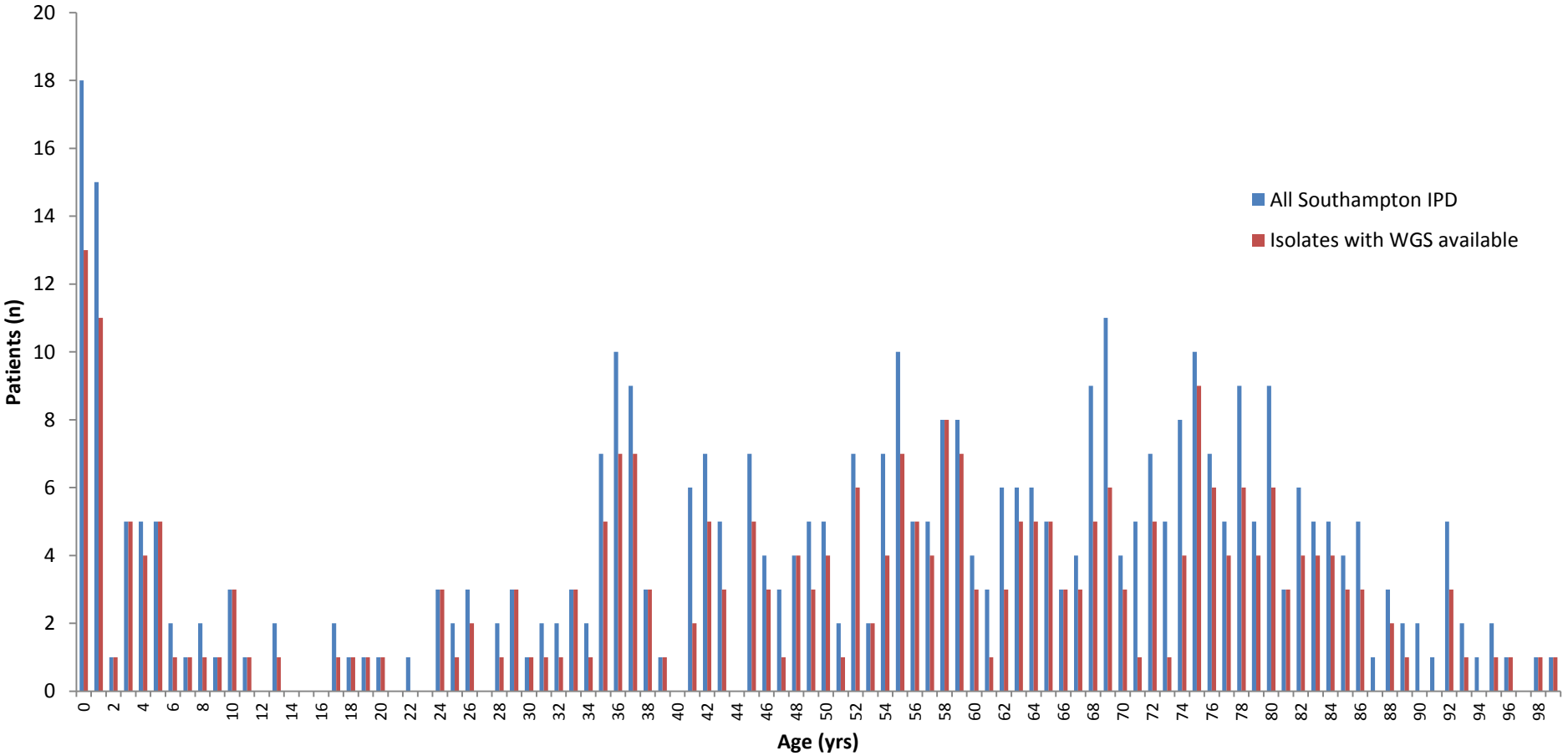
Over 10 years, IPD incidence was highest in the 65+ age category (21 cases per 100,000) and the 0 to 4 years category (18.5 cases per 100,000). IPD incidence was lowest in the 5 to 14 and 15 to 44 years categories. For all ages IPD was estimated to occur in 9.8 per 100,000 people. Table 16 shows these rates and the population data used to calculate them.

Table 16 IPD incidence rate per age category

Age category	No. of (raw) corrected[8] IPD cases	Person-years of observations (10 years)	IPD incidence rate per 100,000 person-years
0 to 4	(44) 44.3	239990	18.5
5 to 14	(17) 17.1	415790	4.1
15 to 44	(77) 77.6	1726000	4.5
45 to 64	(107) 107.8	1002480	10.8
65 +	(156) 157.2	749180	21
All ages	(401) 404	4133440	9.8

[8] Corrected for missing age data (n= 3).

Figure 23 Age distribution of IPD patients



4.3.3 IPD incidence per age category over time

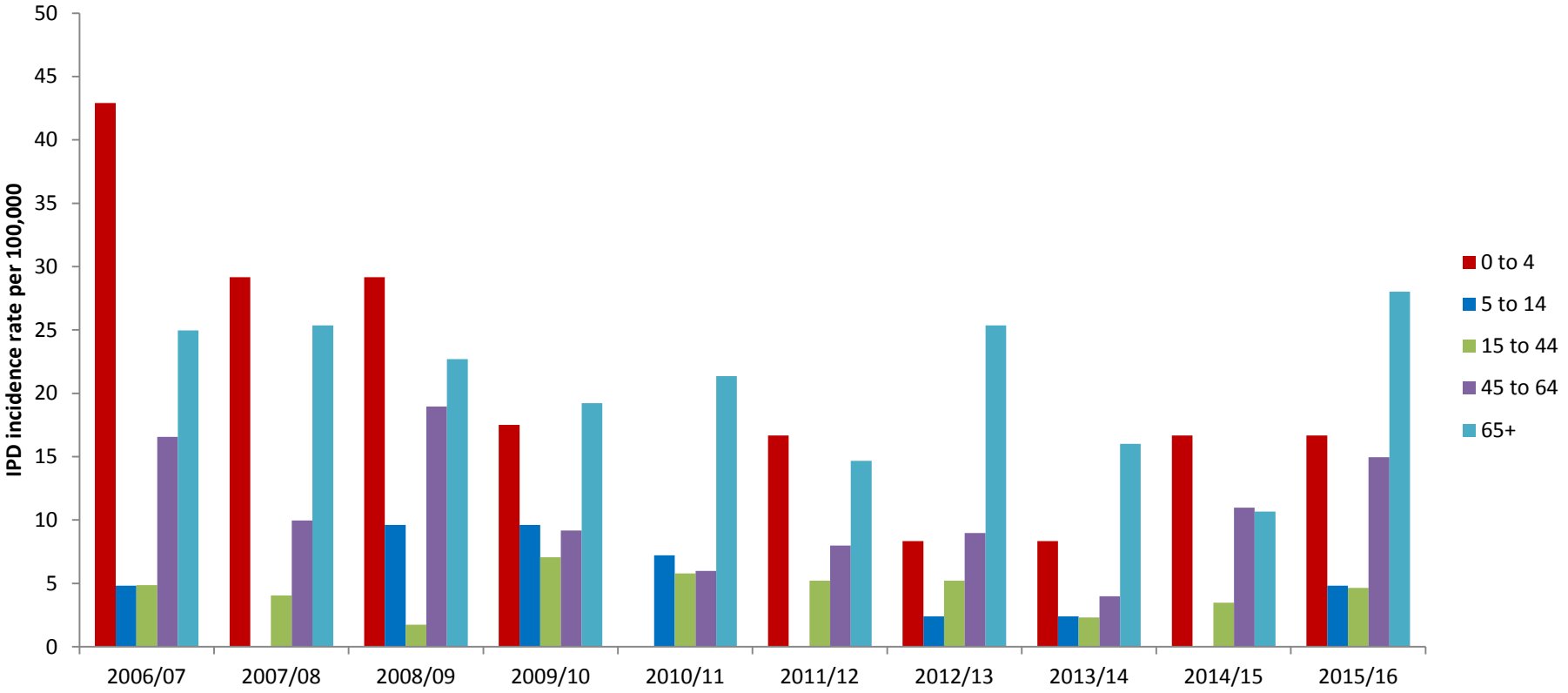
The incidence rate was calculated for each age group for every epidemiological year. The results are presented in Figure 24.

IPD incidence in patients aged 0 to 4 years significantly decreased over the 10 years (p= 0.0004). The highest incidence rate was observed in the first year (2006/07) when 42.9 cases per 100,000, was recorded. The lowest incidence rate was seen in 2010/11 when no case of IPD in this age group was recorded. An increase in the IPD incidence rate has been observed in the years following 2012/13. In 2015/16 the IPD incidence rate was 16.7 cases per 100,000.

For the age categories 5 to 14 years, 15 to 44 years and 45 to 64 years the IPD incidence rate remained statistically stable over the 10 years. Of these three categories, IPD incidence was highest in the 45 to 64 year olds. In the epidemiological years 2010/11 and 2012/13, IPD incidence was higher in this age group than in the 0 to 4 year category.

For patients aged 65+ years, IPD incidence remained statistically stable over the 10 years (p= 0.1). In the first three epidemiological years and in 2011/12, IPD incidence was higher in the 0 to 4 year old than in patients aged 65+. The highest incidence rate was observed in 2015/16 which was recorded to be 28 cases per 100,000.

Figure 24 IPD incidence per age category over time

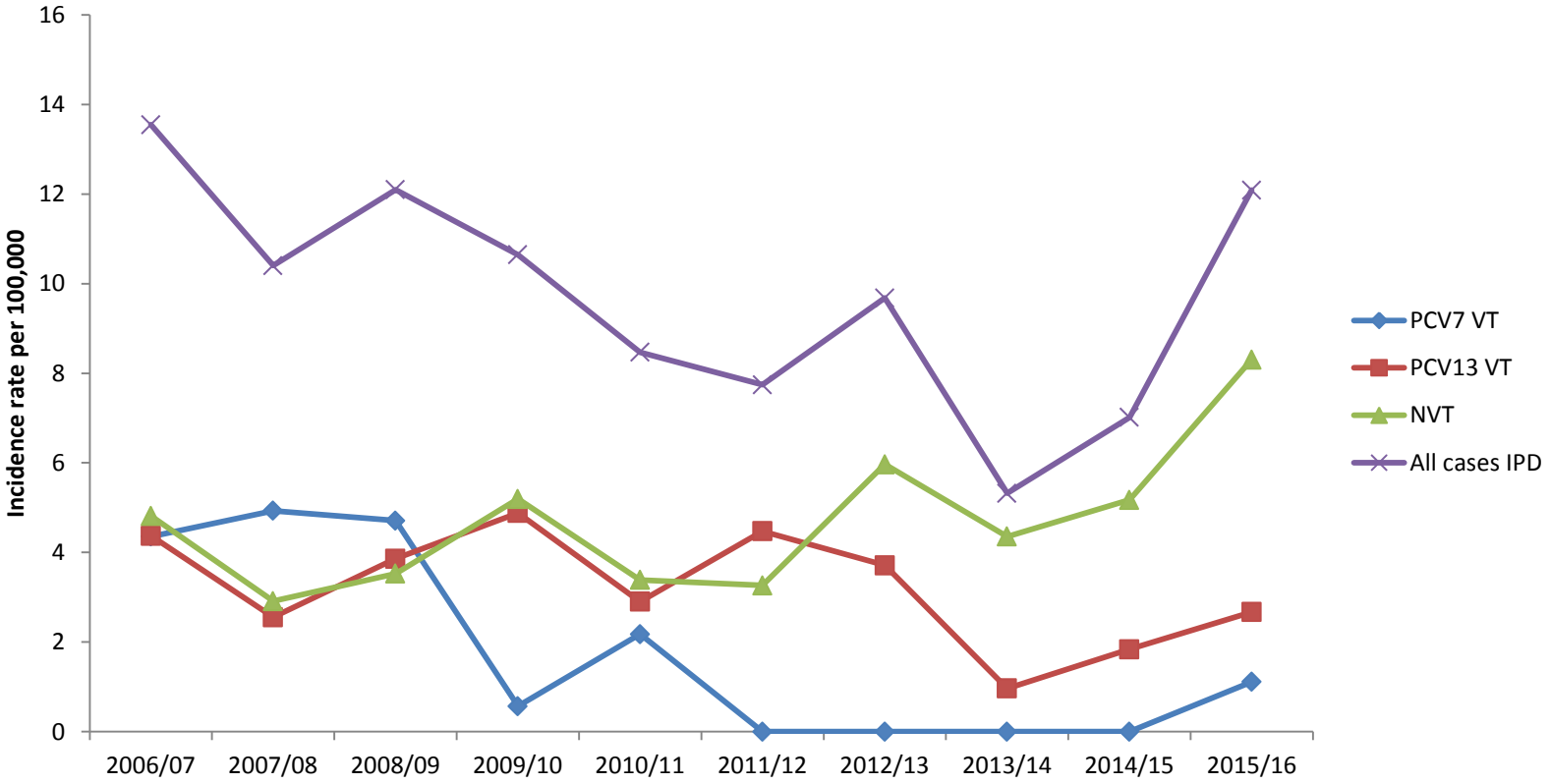


4.3.4 IPD caused by VT serotypes

IPD caused by PCV7 VTs decreased over the 10 years (p= <0.0001). In 2006/07 the incidence rate was 4.4 cases per 100,000 and similar rates were seen in the following two years. PCV7 VT incidence decreased and was not observed at all between 2011/12 and 2014/15. In 2015/16, serotypes 4, 9V and 14 were isolated from cases of IPD. The serotype 4 isolate came from the blood of a 43 year old male who presented in the emergency department of SGH, the 9V was found in the blood of a 68 year old male cancer patient and the serotype 14 was isolated from a 29 year old male emergency department patient.

IPD caused by PCV13 serotypes was also observed to decrease overall (p= 0.02). The highest incidence rate was recorded in 2009/10 and the lowest in 2013/14. Analysis of individual PCV13 serotypes revealed that only serotypes 1 and 7F had a statistically significant incidence decrease over 10 years, p= 0.004 and p= 0.02, respectively. Serotype 7F was also found to have the highest rate of incidence of all serotypes over 10 years (10.7 cases per 100,000). Serotypes 3, 6A and 19A remained at a stable rate over the same time period. Serotype 5 was the only VT serotype to not be isolated in the 10 years. Figure 25 shows the IPD incidence caused by VT and NVT serotypes.

Figure 25 Trends in IPD incidence caused by VT and NVT serotypes



IPD caused by PCV7 VT decreased overall (p=<0.0001) as did IPD caused by PCV13 VTs (p= 0.02). IPD caused by NVT serotypes increased over 10 years (p= 0.004).

4.3.5 IPD caused by NVT serotypes

A total of 22 NVT serotypes were identified over the 10 years and overall IPD caused by NVT serotypes increased during this time (p= 0.004). Serotypes 8, 22F and 12F were the NVTs with the highest rate of IPD cases overall. The overall incidence trend for IPD caused by NVT serotypes is shown in Figure 25.

The highest NVT incidence rate was observed in 2015/16 when they made up 68.7% of the IPD isolates. In the first study year (2006/07) NVT serotypes made up 35.6% of all IPD isolates.

Of the 11 NVT serotypes isolated in 2015/16, four of them, serotypes 8, 22F, 11A and 38 were also observed in the first study year. Serotype 8 is the only serotype that has featured in every year of the study. Four NVT serotypes increased significantly over the 10 years, 23A (p= 0.05), 24F (p= 0.02), 16F (p= 0.04) and 12F (p= <0.0001). 23A was first observed in 2010/11, 24F in 2009/10, 16F in 2013/13 and 12F in 2007/08.

A comparison of the first study year (2006/07) and the last (2015/16) is shown in Figure 26.

Figure 26 Comparison of IPD in first and last study years

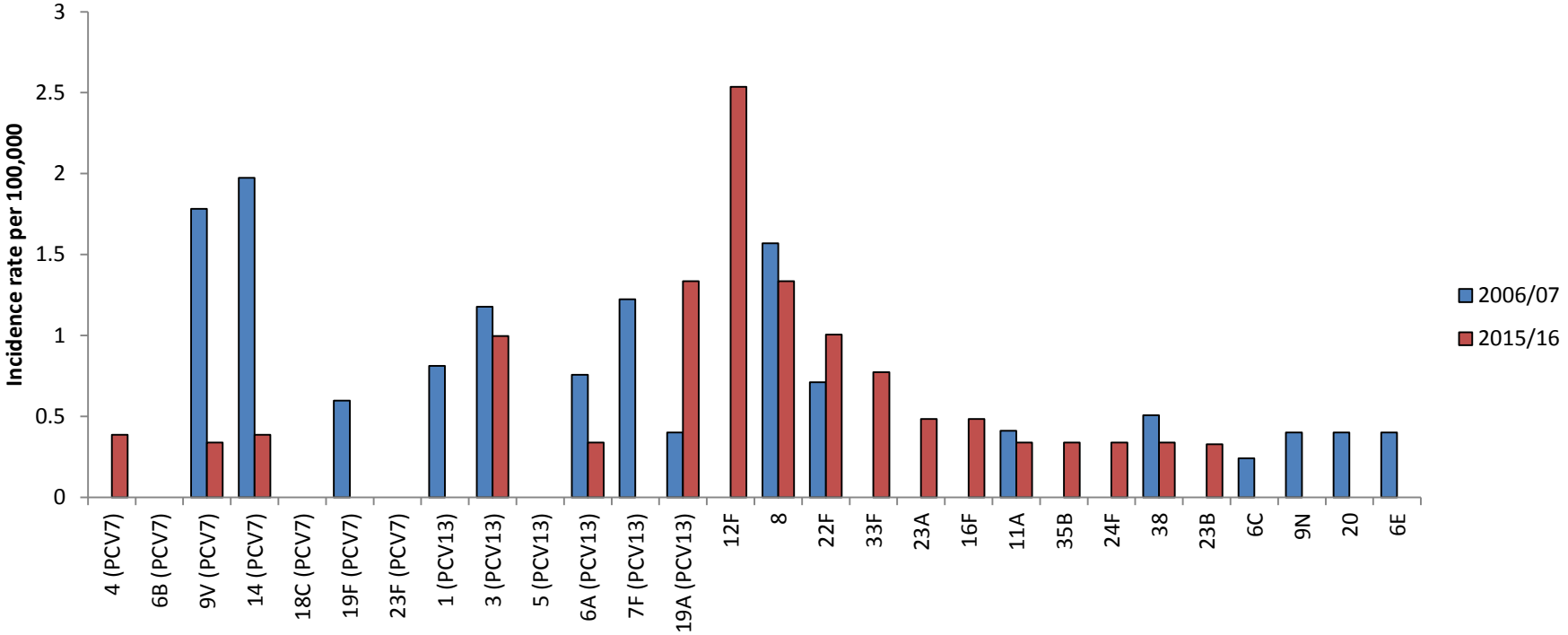


Figure 26 shows the serotype distribution of IPD cases in 2006/07 and 2015/16. NVT serotypes 12F and 8 and PCV13 VT 19A were responsible for the highest incidence rates in 2015/16. Serotypes 8 and 19A were present in the first study year (2006/07) and serotype 12F was first observed in the epidemiological year 2007/08.

4.3.6 Serotypes causing IPD in children 0 to 4 years old

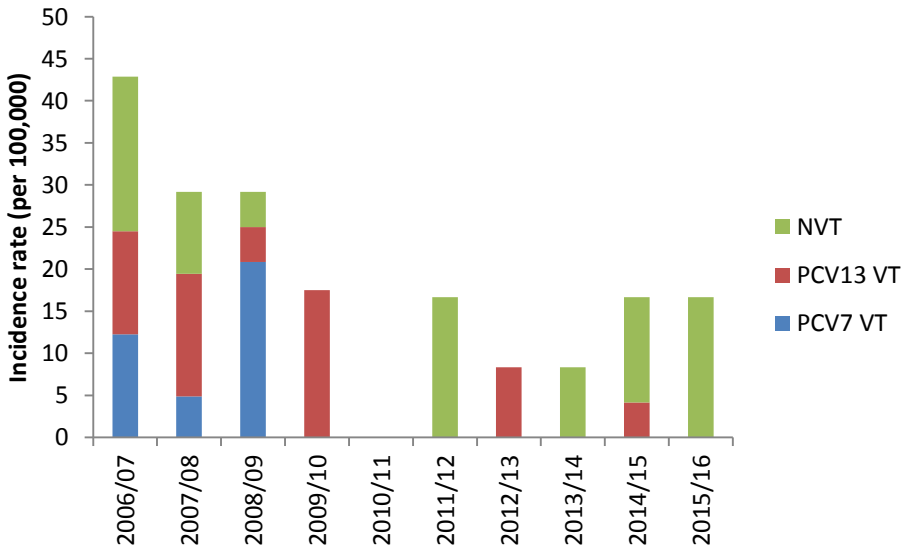
A total of 44.3 (adjusted for missing age data; raw number = 44) children aged 0 to 4 years had confirmed cases of IPD between July 2006 and June 2016.

A total of 18 distinct serotypes were identified from this age group. Of the 13 serotypes covered by PCV13, 11 were found in IPD cases (6B and 5 were not observed). A significant decrease in IPD caused by PCV7 and PCV13 VT serotypes occurred over 10 years with chi squared test for trend results of p= 0.002 and p= 0.04, respectively. The most prevalent VT was serotype 3 which was identified in three epidemiological years; 2006/07, 2007/08 and 2012/13. The latest VT responsible for IPD was 19A which was isolated from a 6 week old male infant in 2014/15.

IPD caused by NVTs remained statistically stable (p= 0.5) over the 10 years. A total of 7 NVT serotypes were observed: 15A, 22F, 15B/C, 23A, 10A, 12F and 8. The most prevalent of these was serotype 22F (n = 7.1 adjusted for missing data, raw number = 6). No NVT were observed to either increase or decrease over the 10 years. Two NVTs were responsible for IPD in 2015/16, 23A and 12F.

Figure 27 shows the distribution of VT and NVT serotypes in the 0 to 4 age group over 10 years.

Figure 27 VT:NVT distribution in the 0 to 4 age group



4.3.7 Serotypes causing IPD in older adult patients

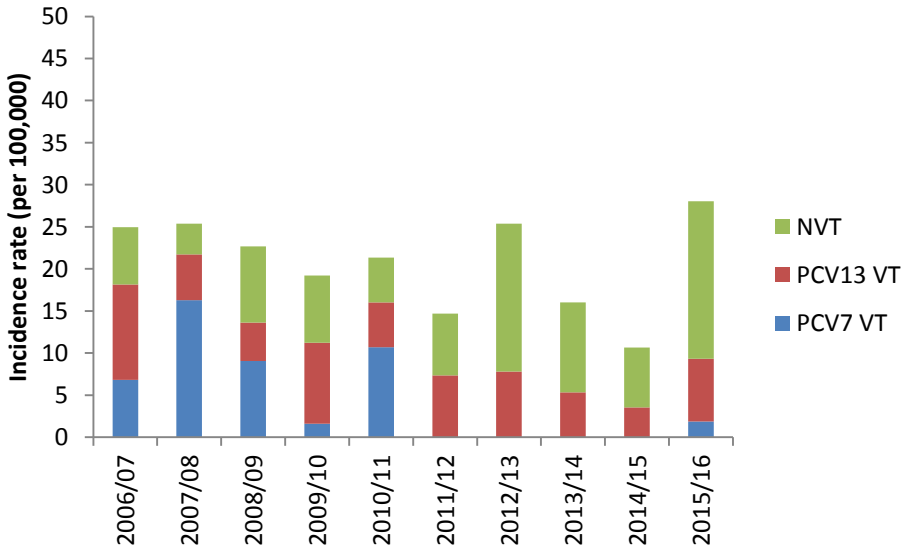
From July 2006 to June 2016 a total of 157.2 (adjusted for missing age data; raw number = 156) older adults aged 65 years and over had IPD.

Of the 13 serotypes in PCV13, 11 were identified from IPD cases with only VTs 18C and 5 unobserved during this time. IPD caused by PC7 VTs decreased significantly (p= <0.0001) while IPD caused by PCV13 VTs remained stable (p= 0.3). The most prevalent PCV13 VT was 19A which was seen in 7 of the 10 epidemiological years. In 2015/16, four VTs were isolated from cases of IPD in older adults; PCV7 VT 9V and PCV13 VTs 3, 6A and 19A.

IPD caused by NVTs increased significantly over 10 years (p= 0.004). A total of 18 serotypes were observed over time and the most prevalent were NVTs 8, 23A and 11A. A significant increase was seen in serotype 12F (p= 0.007) although this was only observed in 2015/16 (n= 2.8 adjusted for missing data, raw number = 2). A significant decrease was seen for 6C (p= 0.02) which was not observed after the epidemiological year 2008/09.

Figure 28 shows the distribution of VT and NVT serotypes in older adults aged 65 years and over.

Figure 28 VT:NVT distribution in the over 65 age group



4.3.8 Invasive potential of individual serotypes

The size of the carriage and the IPD data sets have allowed a broad analysis into the potential impact each serotype may have on disease. Carriage rates were compared to incidence rates to gain an understanding of serotype-specific prevalence and this is shown in Figure 29 and Figure 30. Both the carriage and the IPD data cover the 10 years 2006 to 2016.

Of the VT serotypes, 7F had the highest incidence rate (10.7 cases per 100,000) and a low carriage rate (0.7% over 10 years) relative to other serotypes. VTs 6A and 6B were isolated more frequently from carriage isolates than from IPD cases and serotype 19A was prevalent in both datasets.

Of the NVTs shown in Figure 30, serotype 8 is shown to have a high IPD incidence rate relative to other NVTs and a low carriage rate. Serotypes 15B/C and 11A were more prevalent in the carriage dataset and 22F featured in both IPD and carriage.

Additional to serotype-specific prevalence in carriage and disease, it is useful to examine the potential that each serotype has for invasive disease. This may be presented in the form of odds ratios (OR) which are calculated by comparing the odds of a given serotype appearing in invasive disease to the odds of that serotype being observed in carriage. An OR of >1 would indicate that there is an increased potential of that serotype featuring in disease while an OR of <1 denotes decreased disease potential.

The ORs of each VT serotype were calculated and are shown in Table 17. Of the VTs, serotype 4 had the highest OR (16.21) although other serotypes were more prevalent in IPD. Serotype 7F had a high OR (15.95) and had a higher incidence rate relative to other VTs. In total, 10 out of the 13 VTs had a statistically significant OR. Eight of these had greater odds of featuring in IPD: 4, 14, 9V, 18C, 1, 3, 7F and 19A. Two had greater odds of appearing in carriage, 6B and 6A.

Of the NVTs, serotype 8 had the highest OR (36.43), followed by 12F (17.61). Both serotypes also had high incidence rates relative to other serotypes. A total of 22 NVT serotypes plus non-typeable (NT) pneumococci were observed in both carriage and disease. 11 serotypes were found to have statistically significant ORs, eight of these had greater odds of featuring in carriage than disease; 11A, 6C, 23B, 15B/C, 21, 15A, 35B and 10A. Three serotypes had greater odds of appearing in IPD than carriage; 9N, 12F and 8.

Figure 29 Comparison of carriage rate and incidence rate of VT serotypes

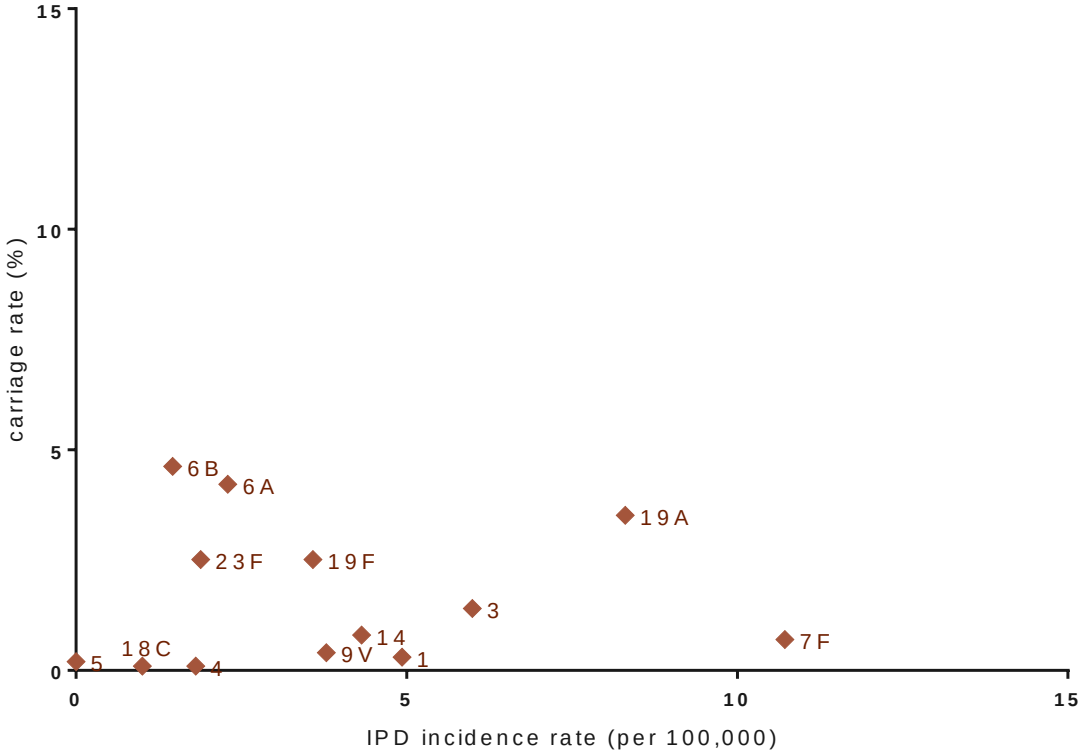


Figure 30 Comparison of carriage rate and incidence rate of NVT serotypes

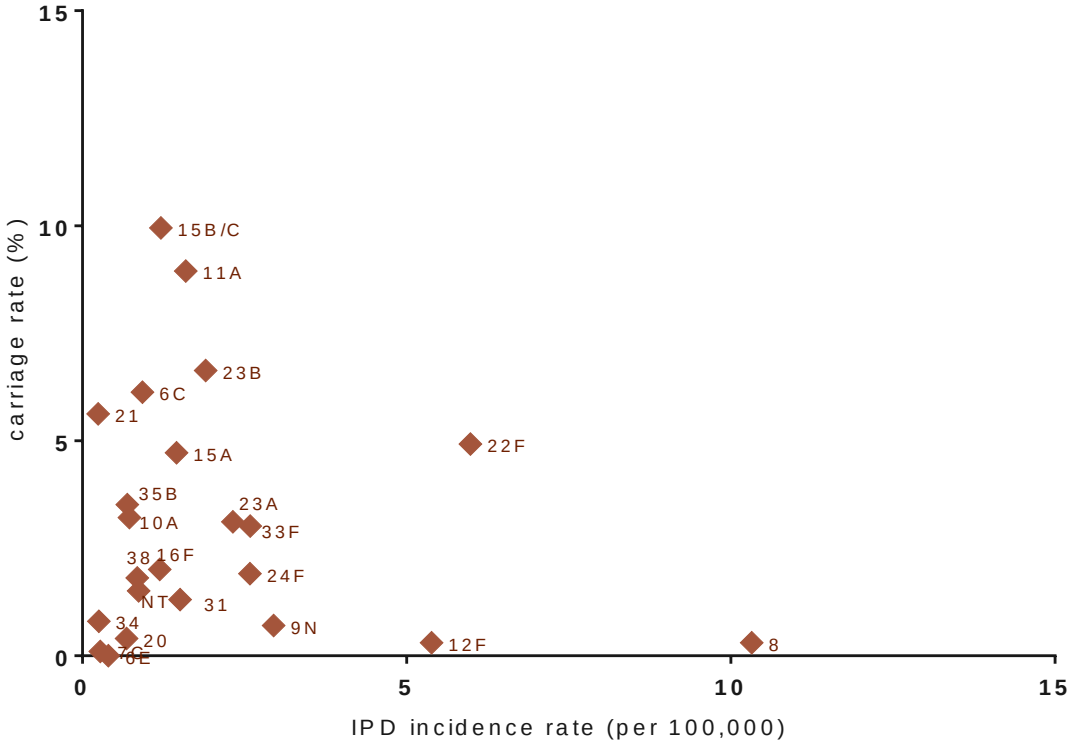


Table 17 Odds ratios to calculate IPD potential of VT serotypes

Serotype	IPD (n)	Carriage (n)	Odds ratio	95% CI	p-value
4 (PCV7)	7	1	16.21	2 to 132.2	0.009
6B (PCV7)	6	46	0.29	0.1 to 0.7	0.004
9V (PCV7)	16	4	9.45	3.1 to 28.4	0.0001
14 (PCV7)	18	8	5.32	2.3 to 12.3	0.0001
18C (PCV7)	4	1	9.19	1 to 82.5	0.05
19F (PCV7)	15	25	1.38	0.7 to 2.6	0.3
23F (PCV7)	8	25	0.72	0.3 to 1.6	0.4
1 (PCV13)	20	3	15.93	4.7 to 53.9	<0.0001
3 (PCV13)	25	14	4.27	2.2 to 8.3	<0.0001
5 (PCV13)	0	2	0.45	0.02 to 9.5	0.6
6A (PCV13)	9	42	0.48	0.2 to 1	0.05
7F (PCV13)	44	7	15.95	7.1 to 35.7	<0.0001
19A (PCV13)	34	35	2.32	1.4 to 3.8	0.0007

Table 18 Odds ratios to calculate IPD potential of NVT serotypes

Serotype	IPD (n)	Carriage (n)	Odds ratio	95% CI	p-value
11A	7	89	0.16	0.08 to 0.4	<0.0001
6C	4	61	0.14	0.05 to 0.4	0.0002
23B	8	66	0.26	0.1 to 0.5	0.0004
15B/C	5	99	0.10	0.04 to 0.3	<0.0001
21	1	55	0.04	0.005 to 0.3	0.001
15A	6	47	0.28	0.1 to 0.7	0.004
22F	25	49	1.17	0.7 to 1.9	0.5
23A	10	31	0.73	0.4 to 1.5	0.4
35B	3	35	0.19	0.06 to 0.6	0.006
33F	11	30	0.83	0.4 to 1.7	0.6
10A	3	32	0.21	0.06 to 0.7	0.01
24F	11	19	1.33	0.6 to 2.8	0.5
16F	5	20	0.56	0.2 to 1.5	0.3
38	4	18	0.50	0.2 to 1.5	0.2
31	6	13	1.05	0.4 to 2.8	0.9
NT	4	15	0.60	0.2 to 1.8	0.4
34	1	8	0.28	0.04 to 2.3	0.2
9N	12	7	3.99	1.6 to 10.2	0.004
20	3	4	1.71	0.4 to 7.7	0.5
12F	22	3	17.61	5.2 to 59.2	<0.0001
8	43	3	36.43	11.2 to 118.2	<0.0001
7C	1	1	2.28	0.1 to 36.6	0.6
6E	2	0	11.40	0.5 to 238.8	0.1

4.4 Discussion

Overall IPD has decreased in the 10 years between 2006 and 2016 although examination of annual data has shown that this decrease is only true for the first seven years. In 2013/14 the incidence rate was the lowest recorded for this study period and incidence has risen in the years following.

IPD surveillance data for England and Wales collated by PHE has also reported an overall decrease in IPD incidence. For 2008/10 the incidence rate for England and Wales was recorded as 10.13 cases per 100,000 (Ladhani *et al.*, 2018). In the Southampton area for the same time period, incidence was recorded as 12.3 cases per 100,000 in 2008/09 and 10.6 cases per 100,000 in 2009/10.

For the epidemiological year 2013/14 PHE reported that IPD in England and Wales had decreased to 6.85 cases per 100,000 (Waight *et al.*, 2015) and a similar decrease was observed in the Southampton data where IPD incidence had fallen to 5.8 cases per 100,000.

A rise in IPD cases was observed both nationally and locally in the period following 2013/14. The latest PHE data has reported an incidence rate of 9.87 for England and Wales. A higher incidence rate was found for the Southampton area with 12 cases per 100,000 observed in 2015/16. A follow-up analysis has shown that this increase has continued and IPD incidence was 14 cases per 100,000 in the Southampton area in 2016/17 (data for 2016/17 not shown in this thesis). This is a higher incidence rate than the first study year of 2006/07.

Contrary to the PHE data which shows that IPD incidence has only increased in the 65 and over age category, the Southampton data shows that IPD has also increased in children aged 0 to 4 years from 4.2 cases per 100,000 in 2013/14 to 16.7 cases per 100,000 in 2015/16.

The serotypes responsible for the overall increase since 2013/14 are primarily NVTs which made up 69.8% of IPD isolates. PCV7 VTs made up 3.2% of IPD cases in the same three years and the additional six VTs found in PCV13 were responsible for 27% of IPD cases. In 2015/16 NVTs made up 69.2% of IPD cases, PCV7 VTs made up 9.2% of IPD cases and the additional six VTs in PCV13 constituted 21.7% of IPD.

The reappearance of PCV7 VTs in 2015/16 is of concern. They were last seen in 2010/11 and went unobserved in IPD for four years. Three serotypes were isolated in 2015/16, two of them, VTs 4 and 14, were found in patients of the 15

to 44 age category and one, VT 9V was isolated from a patient aged over 65 years. PCV7 VTs have not appeared in IPD in the 0 to 4 age group since 2008/09.

PCV13 VTs decreased over the period of the study. The serotypes responsible for this overall decrease were 1 and 7F, other serotypes, 3, 6A and 19A remained statistically stable. A significant decrease in PCV13 VTs was only observed in one age group, 0 to 4 years. In all other age groups IPD caused by the additional six PCV13 VTs remained statistically stable.

The PHE IPD data has previously been used to examine the impact of PCV usage in young children on IPD in non-PCV recipients (Miller *et al.*, 2011, Waight *et al.*, 2015, Ladhani *et al.*, 2018). The Southampton data supports the reduction in PCV7 VTs in non-recipients however the effect of PCV13 on older adults has not been seen to the same extent. Whilst the overall incidence rate was similar for the 65 years and over age group (28.9 per 100,000 from PHE's 2016/17 data and 28 per 100,000 in the Southampton area 2015/16) the distribution of VTs and NVTs were different. For England and Wales PCV13 VTs were responsible for 5.5 cases per 100,000 in people age 65 years and over and in Southampton PCV13 VTs were responsible for 7.5 cases per 100,000 in the same age group.

Three NVT serotypes were found to have significant odds of appearing in IPD, 8, 12F and 9N. Serotypes 8 and 9N have remained stable over the 10 years however there has been a significant increase in serotype 12F. An increase in IPD caused by 12F was seen in every age group except the 5 to 14 year olds. By 2015/16 it was responsible for 20.1% of IPD cases in the Southampton area and had an incidence rate of 2.5 cases per 100,000. It is rarely carried and has only been isolated three times in the 10 years of the carriage study.

Examination of IPD data in a variety of regions is important for public health. Regional trends caused by demographic differences may not appear in larger data sets such as the one compiled by PHE.

Chapter 5 Molecular epidemiology of pneumococcal carriage and disease: 2006 - 2016

5.1 Introduction

The availability of both carriage and IPD data that is matched for location and time provides a valuable opportunity to examine pneumococcal epidemiology. Whilst examination of serotype distribution is useful for assessing disease potential and vaccine impact it does not provide information pertaining to the relatedness of carriage and disease or explain how the pneumococcus adapts to the impact of clinical interventions such as vaccines.

A lack of community wide surveillance studies mean that little is known about pneumococcal transmission. Evidence that pneumococcal carriage in children is related to disease in older adults has been presented in studies within familial settings (Dowling *et al.*, 1971, Owen Hendley *et al.*, 1975) and from seasonal disease rates (Walter *et al.*, 2009). However, other studies have been unable to find evidence of transmission from children to older adults (Regev-Yochay *et al.*, 2004).

The availability of WGS data has provided an opportunity to characterise the pneumococcal isolates at a molecular level. Targeted identification of specific genomic regions allows examination of pneumococcal clonality through the application MLST. Whilst the capsule remains the key determinant in disease potential, much remains unknown regarding the influence that clonal types have on disease. An understanding of clonal distribution and how selective pressure from vaccines has affected this may provide an insight into the impact that pneumococcal clones have on disease.

Genomic data also allows the evolutionary relationships between isolates to be examined. By identifying single nucleotide polymorphisms (SNPs) in genomic sequences the phylogeny of pneumococcal isolates in carriage and disease may be compiled and investigated for genetic relatedness.

5.1.1 Main aims

- To characterise both the carriage and the IPD pneumococcal populations in terms of clonal type.
- To identify any changes to ST associations with serotype.
- To examine molecular relatedness of pneumococcal carriage and disease isolates
- To explore the impact that PCV13 has had on pneumococcal clones.
- To study the phylogeny of carriage and disease isolates

5.2 Methods

Multi locus sequence type (MLST) work was carried out with SRST2 to map FASTQ files against an MLST database which held sequences of seven pneumococcal 'housekeeping' genes: *aroE*; *gdh*; *gki*; *recP*; *spi*; *xpt* and *ddl* (see methods 2.5.2). Clonal complexes were defined as sharing six out of the seven allelic sequences and were calculated and visualised using eBURST and Phylovz (methods 2.5.3).

Identification of SNPs was performed using parSNP which is part of the Harvest suite. Resultant trees were then visualised using Microreact (methods 2.5.4).

Statistical analyses were conducted using Graphpad Prism v7.03 (GraphPad Software, La Jolla California USA, www.graphpad.com).

5.3 Results

5.3.1 Pneumococcal sequence types

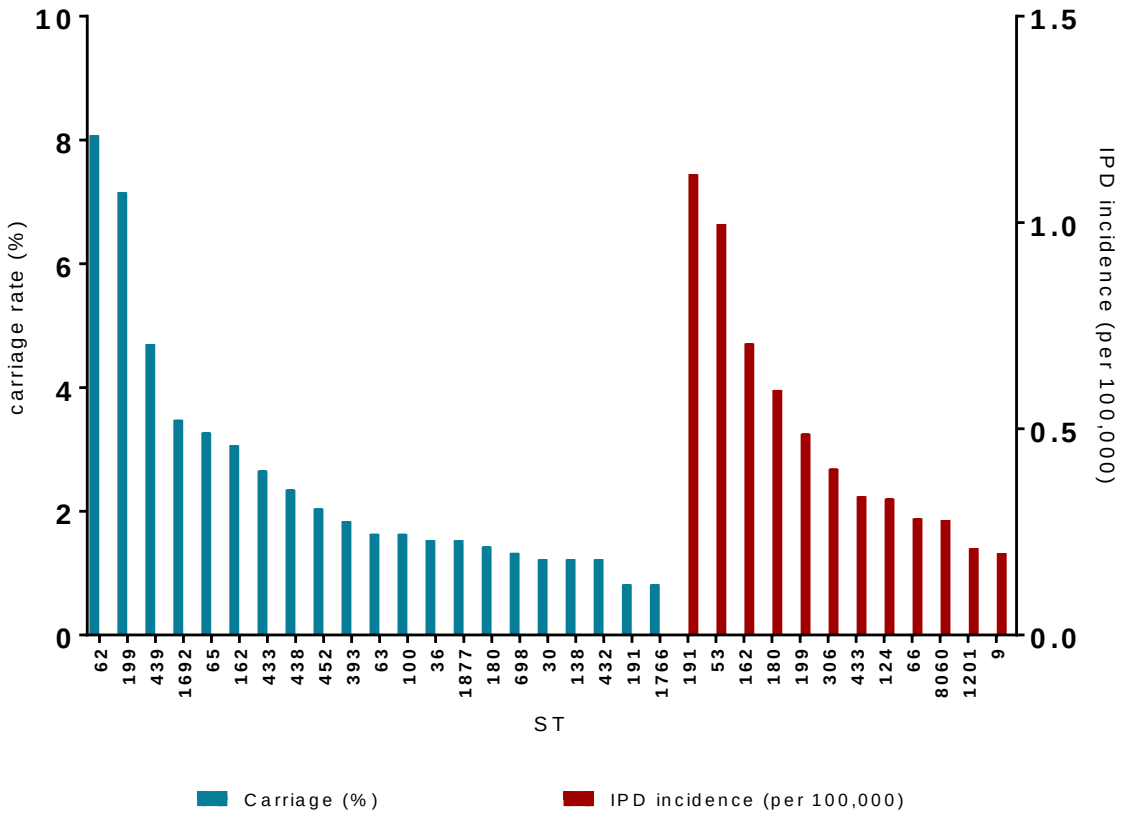
A total of 978 carriage isolates were successfully sequence typed over 10 years. 177 distinct sequence types (STs) were identified and 97 of these appeared only once. 14 STs constituted 50% of all carriage isolates, the most prevalent being ST62 (n= 79) which is associated with serotype 11A and ST199 (n= 70) which was associated with three serotypes, 19A (n= 26), 15B/C (n= 43) and 22F (n= 1). Three STs were observed in every year of the study: ST199, ST62 and ST439.

A total of 286 IPD isolates were successfully sequence typed and 82 distinct STs were identified. Of these 39 STs appeared only once in 10 years. 11 STs constituted 53.5% of all IPD isolates and the most prevalent of these were ST53 (n= 26) which was associated with serotype 8, ST191 (n= 26) associated with 7F and ST162 (n= 19) which was associated with three serotypes, 9V (n= 9), 19F (n=

5) and 24F (n= 5). ST191 had a higher overall incidence rate when the dataset was adjusted for missing data per age and year.

A comparison of carriage and IPD isolates found that 51 STs appeared in both datasets. The majority of IPD isolates (83.3%) had a ST that was also found in carriage. The most prevalent STs in carriage and IPD are displayed on Figure 31. Five STs were observed to feature more than eight times in both carriage and IPD; 199, 162, 433, 180 and 191. Of the 31 STs that featured in IPD but not carriage, ST218 was the most prevalent and was observed in 4 isolates over 10 years. ST218 was associated with serotype 12F.

Figure 31 Prevalent STs in carriage and IPD



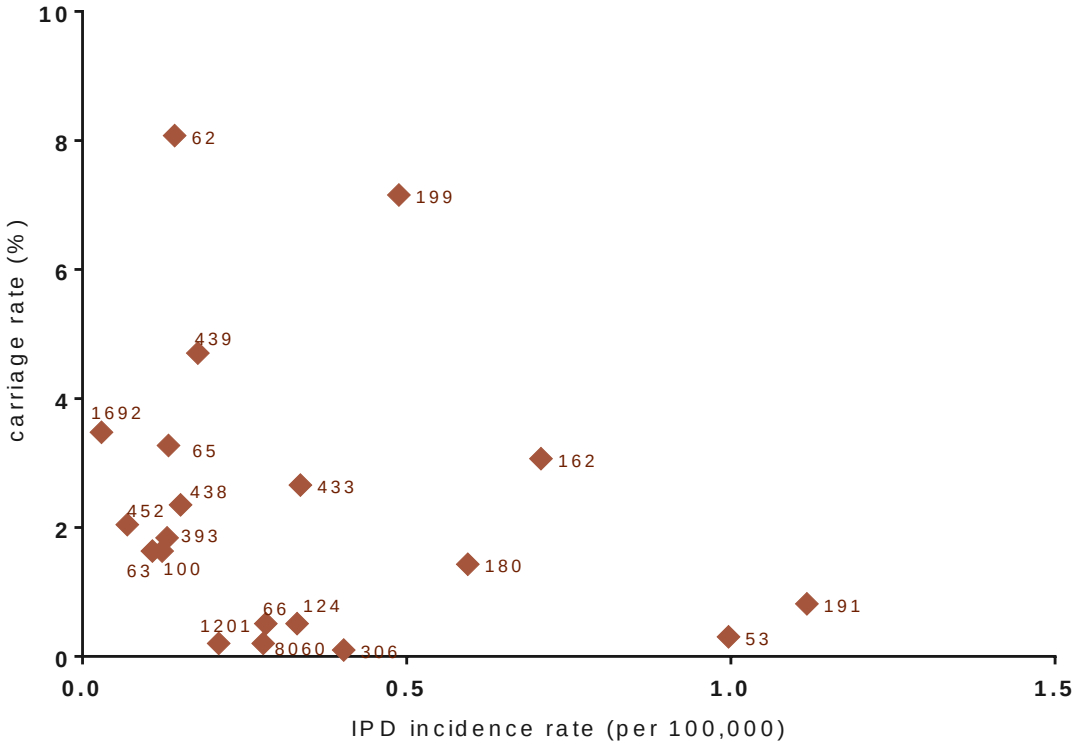
5.3.2 ST prevalence in carriage and IPD

The 51 STs that appeared in carriage and IPD were examined. Figure 32 compares the carriage rate of the most prevalent STs in carriage with the incidence rate of the most prevalent STs in IPD. ST62 was observed to have a high carriage rate relative to other STs and a lower IPD incidence rate. In both carriage and IPD ST62 has been associated only with serotype 11A which was found to have an OR of 0.16.

ST191 which had a relatively higher incidence rate than carriage rate has been primarily associated with serotype 7F although a single carriage isolate was typed as serotype 15A ST191 in 2011/12.

STs 199 and 162 were prevalent in both carriage and IPD and found to be associated with multiple serotypes. ST199 was isolated 70 times in carriage and associated with serotypes 19A (n= 26), 15B/C (n= 43) and 22F (n= 1). In IPD ST199 was only associated with serotype 19A. ST162 was isolated 30 times in carriage and associated with serotypes 9V (n= 3), 19A (n= 1), 19F (n= 9), 15B/C (n= 5), 24F (n= 10), 24B (n= 1) and a non-typeable pneumococci (n= 1). In IPD it was associated with serotypes 9V (n= 3) and 19F (n= 4) and 24F (n= 2).

Figure 32 Comparison of IPD and Carriage STs



5.3.3 ST prevalence over time

Carriage isolates were examined for changes in prevalence between 2006/07 and 2015/16. A chi square test for trend revealed seven STs that had decreased significantly over the 10 years. Six of the STs were primarily associated with VT serotypes and one ST (ST1692) was associated with NVT 6C. They are shown in Figure 33. Only ST62 which is associated with serotype 11A was shown to increase significantly in carriage over the 10 years (p= 0.02).

The smaller number of IPD STs each year meant that statistical analysis had less power. Overall decreases were noted in four STs which are shown in Figure 34. ST191 which was primarily associated with VT serotype 7F peaked in incidence in 2011/12 however it has not been isolated from IPD cases since 2012/13. Increases of individual STs in IPD were noted for ST8060 which is associated with 12F and ST438 which is the primary ST of serotype 23A.

Figure 33 STs observed to decrease in carriage

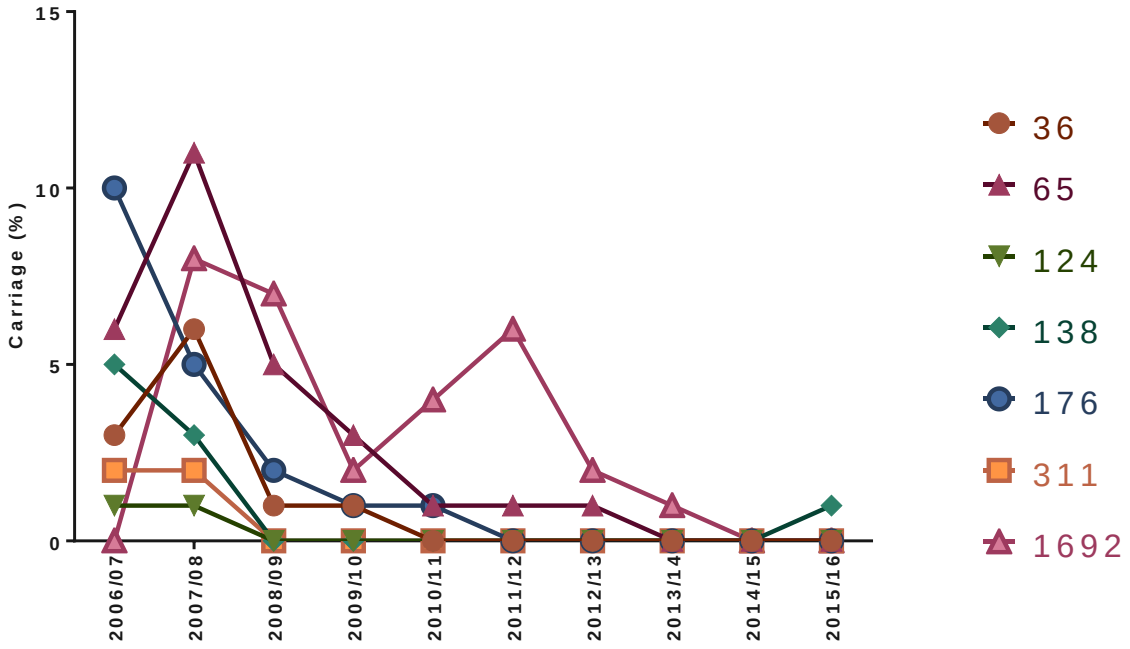
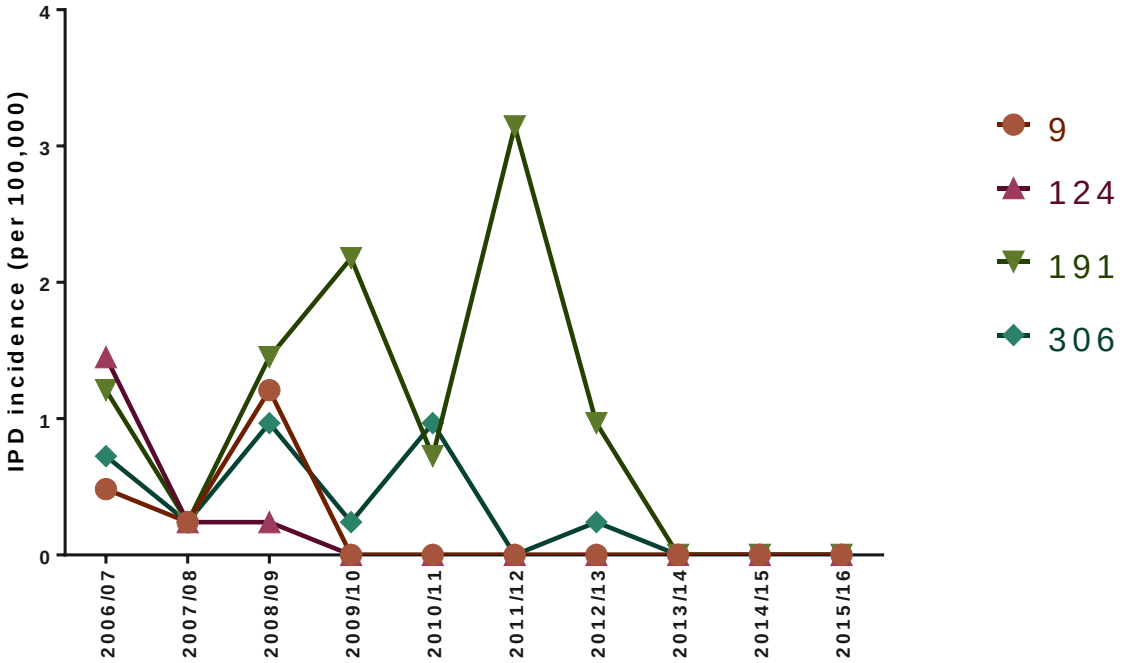


Figure 34 STs observed to decrease in IPD

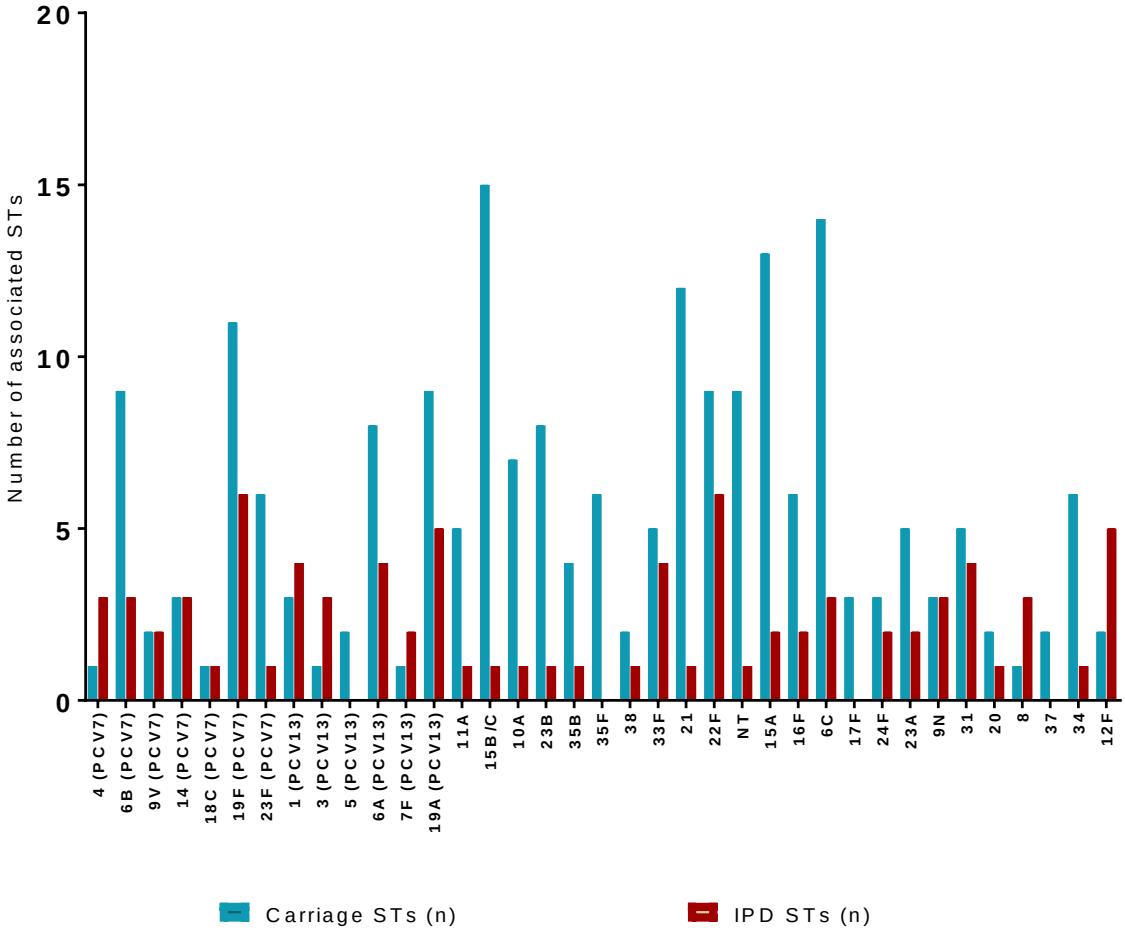


5.3.4 Clonal diversity

The number of STs associated with each serotype found in carriage and IPD was calculated. Amongst the carriage isolates, serotype 15B/C had the highest degree of clonal diversity being observed with 15 distinct STs. Serotypes 22F and 19F had the highest number of associated STs amongst the IPD isolates. Both were observed six times each over the 10 years.

Figure 35 shows each serotype found in carriage and IPD and the number of STs associated with each of them.

Figure 35 Number of STs associated with individual serotypes in carriage and IPD

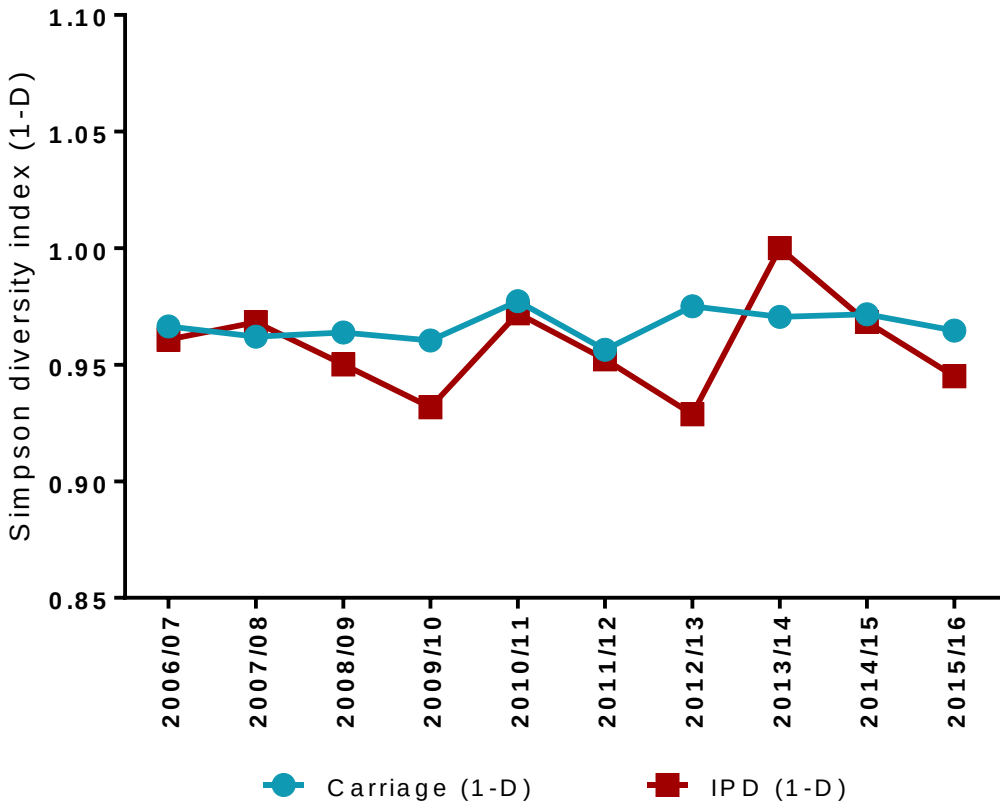


5.3.5 Changes in clonal diversity

The number of STs observed in each of the 10 years was recorded and the Simpson diversity index (1-D) was used to quantify the number of STs observed and their relative abundance over time. Figure 36 shows 1-D values for each of the years. For this analysis only the observed STs were used for each year and the IPD numbers were not adjusted for missing data.

The diversity values for carriage and IPD were analysed with the D'Agostino and Pearson omnibus normality test which estimates if a range of values fall within expected levels of normal variation. The carriage isolates produced a K^2 value of 0.4 (p= 0.8) and the IPD isolates had a K^2 value of 1.1 (p= 0.6) meaning that the diversity values in both datasets did not deviate from a statistically normal range.

Figure 36 ST diversity over 10years

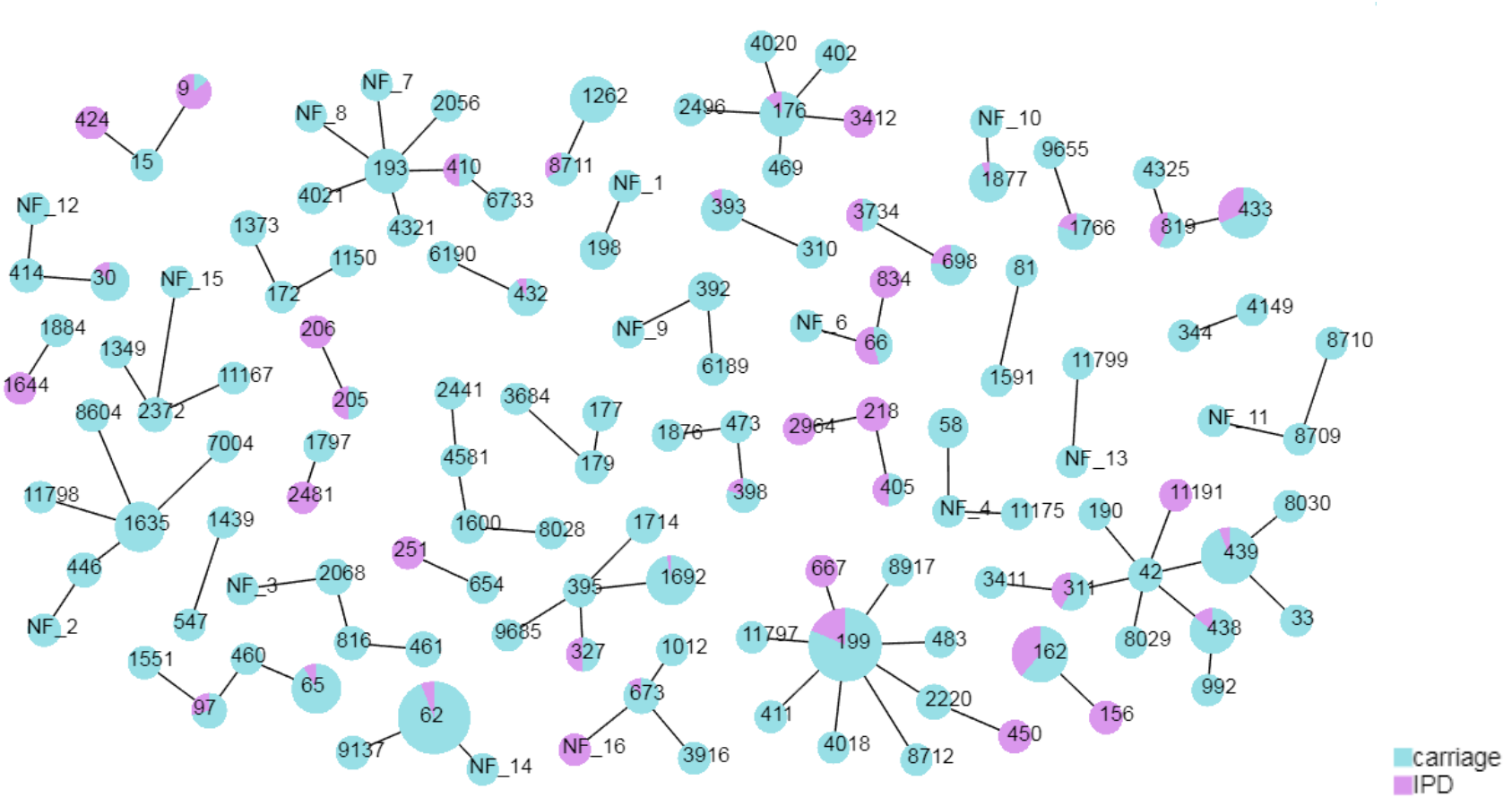


5.3.6 Clonal distribution of carriage and IPD isolates

A comparative analysis of the IPD and carriage isolates was performed. MLST profiles from 1265 isolates (979 from paediatric carriage and 286 isolates from IPD) were used in an eBURST analysis. 208 distinct STs were identified and these were separated into 39 clonal complexes (CC) and 74 singleton STs, i.e. those isolates with five or fewer identical allelic profiles. Figure 37 shows a snapshot of the 39 CCs.

The CCs with greatest number of isolates were CC42 and CC199. CC42 had a predicted founder of ST42 and comprised of 11 STs and a total of 103 isolates. Three of the STs, 311, 438 and 439 were found in both IPD and carriage isolates. One ST, 11191 was observed in IPD only. CC199 had a predicted founder of ST199 and comprised of 104 isolates separated in 10 distinct STs.

Figure 37 Clonal complexes identified in IPD and carriage



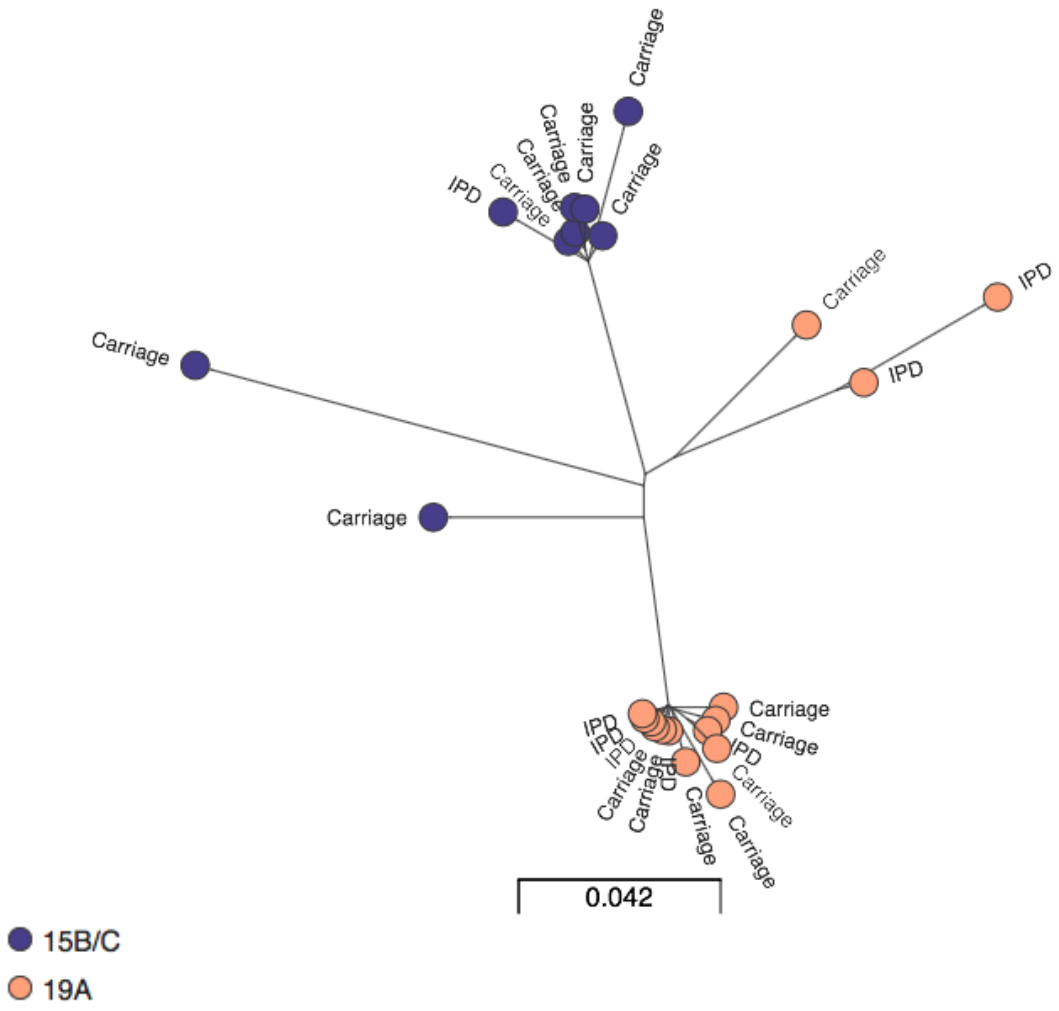
The nucleotide sequences for alleles of seven housekeeping genes were used to assign a ST. STs with six out of seven identical alleles form clonal complexes and are indicated by a line. Frequency of observation is indicated by circle size. NF = not found in database.

5.3.7 Analysis of invasive and carried isolates of ST199

The prevalence of ST199 in carriage and disease plus its inclusion in one of the larger CCs singled it out for further investigation.

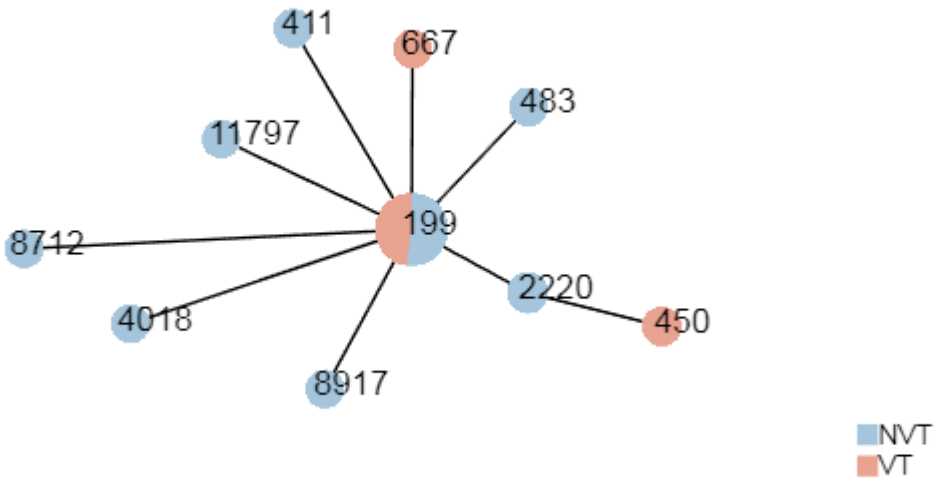
The genomes of ST199 isolates were investigated by combining isolates from IPD and carriage to form a core genome SNP tree. Figure 38 shows the resultant tree. The tree topology showed no distinct clustering that would indicate a separation of evolutionary paths. Instead isolates from both IPD and carriage were well dispersed across the tree.

Figure 38 Core genome SNP tree of ST199 isolates



CC199 was further assessed to identify any changes to serotype associations over the 10 years. Analysis using phyloviz showed that ST199 was observed with both VT and NVT serotypes over the 10 years. The distribution of VT and NVT serotypes is shown in Figure 39.

Figure 39 VT and NVT associations of CC199



VT and NVT associations of carried ST199 isolates over time was investigated and is shown in Figure 40. For the earlier years of the carriage study ST199 was primarily associated with VT serotypes however a possible capsular switch event occurred and the later study period shows that ST199 was primarily associated with NVT serotypes.

In IPD, ST199 is primarily associated with VT serotypes for the entire study period and there is no evidence of capsular switch occurring. Figure 41 shows the VT and NVT associations of ST199 in IPD over 10 years.

Figure 40 VT and NVT associations of carried ST199 over time

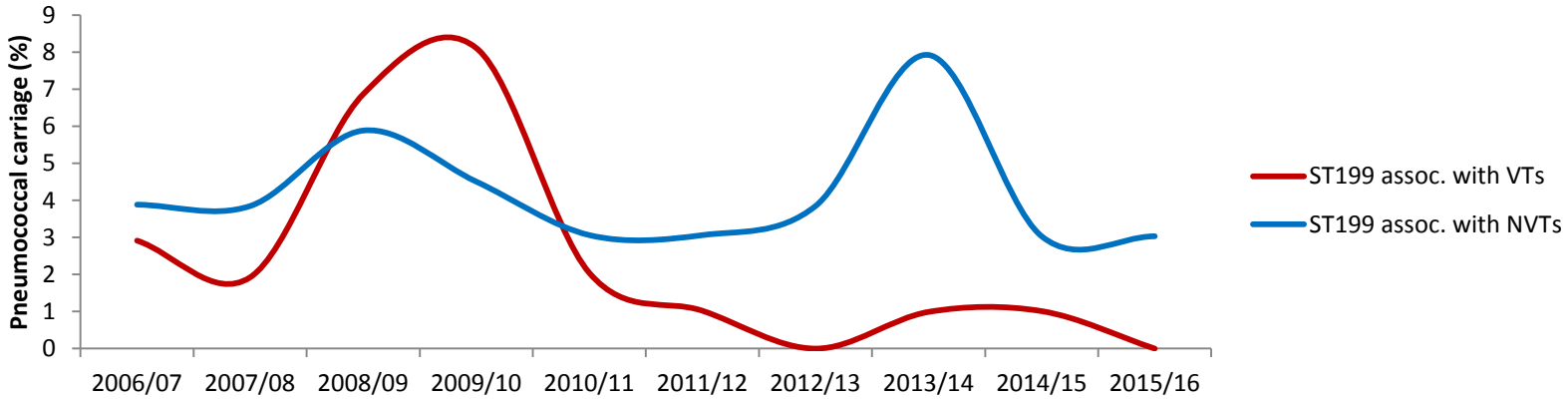
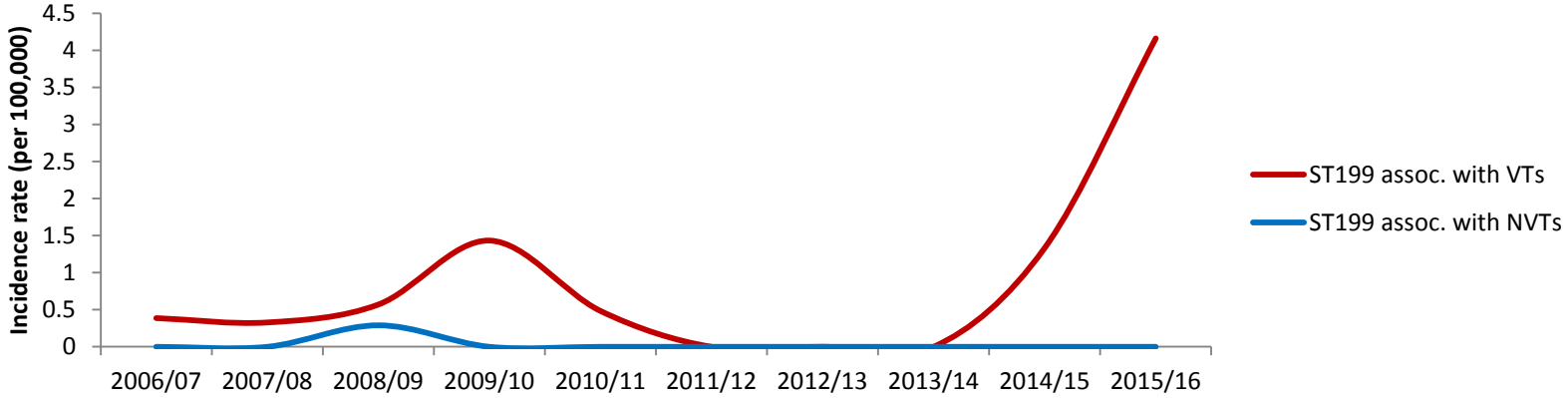


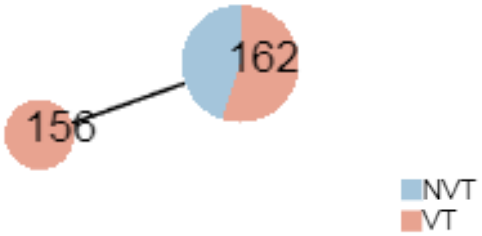
Figure 41 VT and NVT associations of IPD ST199 over time



5.3.8 Analysis of invasive and carried isolates of ST162

The prevalence of ST162 and its association with multiple serotypes in carriage and IPD led to further investigations.

Amongst the carriage isolates, ST162 was found in a CC with only one other ST (ST156). ST156 has only been observed in VT serotypes and ST162 has been associated with both VT and NVT serotypes in carriage and IPD
Figure 42 VT and NVT associations of ST162

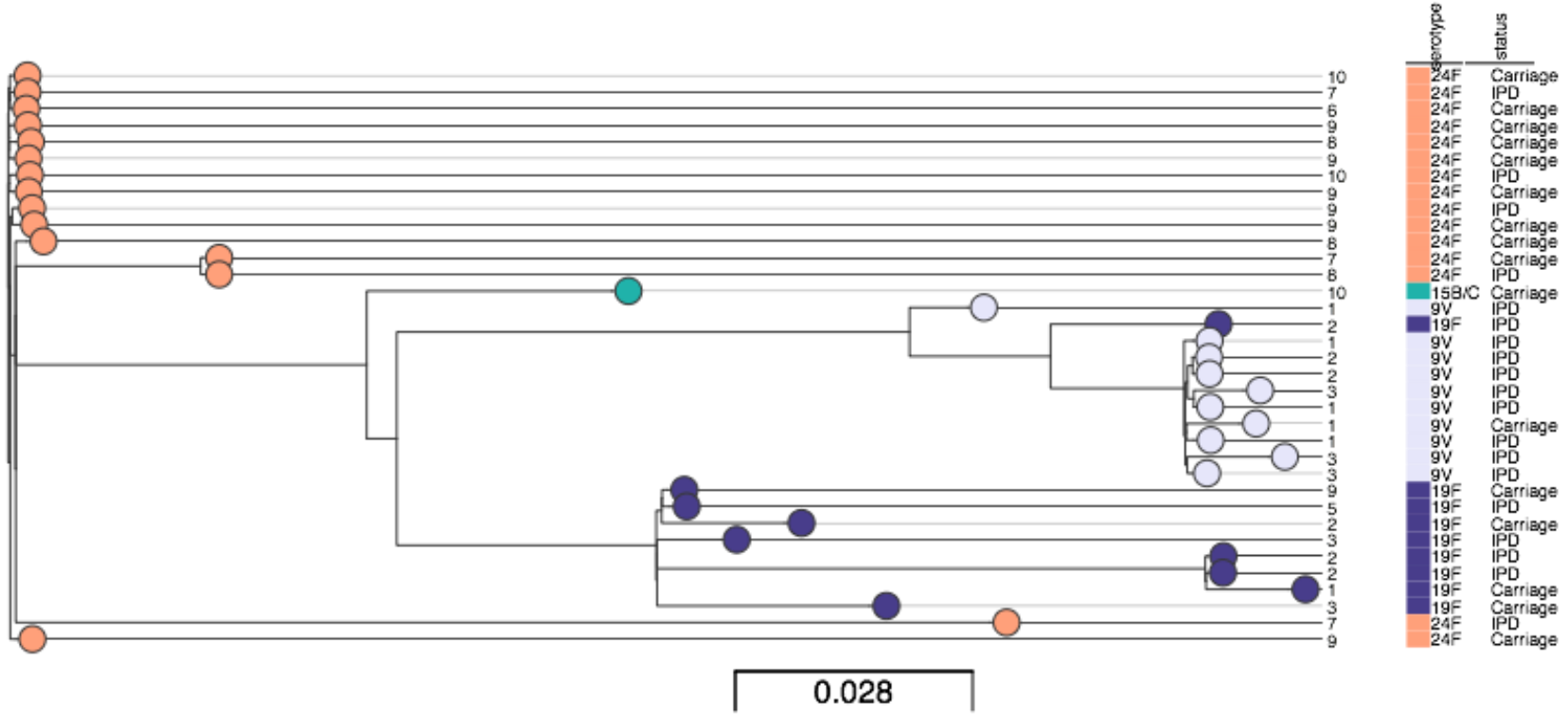


The genomes of ST162 isolates were investigated by combining isolates from IPD and carriage to form a core genome SNP tree. Figure 43shows the resultant tree. Isolates formed clades according to serotype and the carriage and IPD isolates were dispersed within those clades. Four serotypes were found in the carriage isolates and three in the IPD isolates

VT and NVT serotypes associated with ST162 carriage isolates were investigated to see if distribution altered over time. The results are shown in Figure 44. For the earlier years of the carriage study ST162 was primarily associated with VT serotypes and for the later study period ST162 was primarily associated with NVT serotypes.

The ST162 isolates in the IPD dataset displayed a similar shift from VT associations in the earlier study period to NVT in the later IPD. The primary VT serotypes found in both carriage and IPD were 9V and 19F and the NVT was 24F. The ST162 VT and NVT associations for the IPD data are shown in Figure 45.

Figure 43 Core genome SNP tree of ST162 isolates



The numbers at the end of the tree labels indicate the study year with 1 being 2006/07 and 10 being 2015/16.

Figure 44 VT and NVT associations of carried isolates of ST162 over time

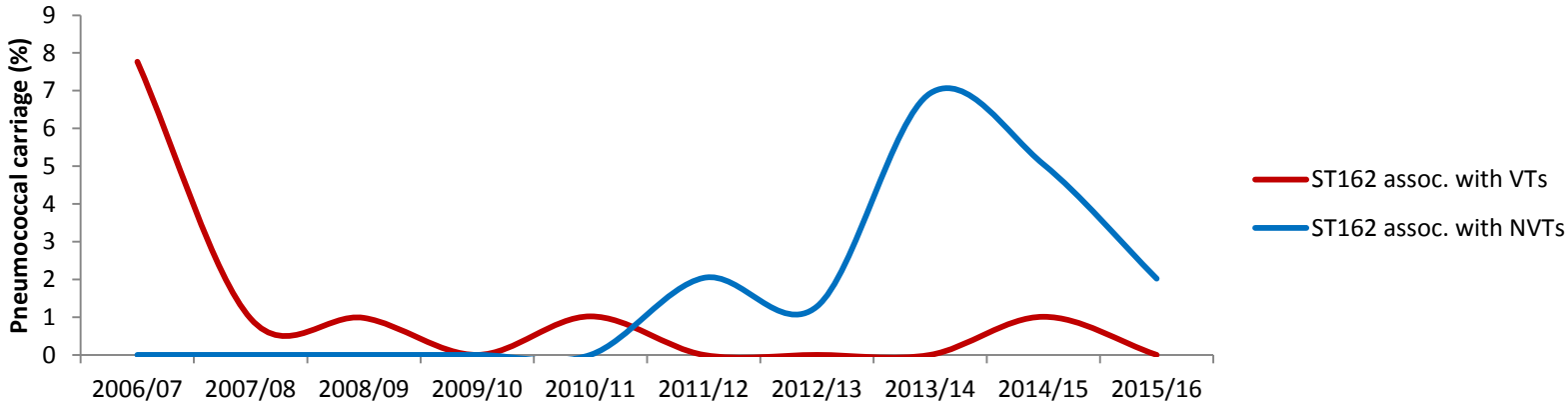
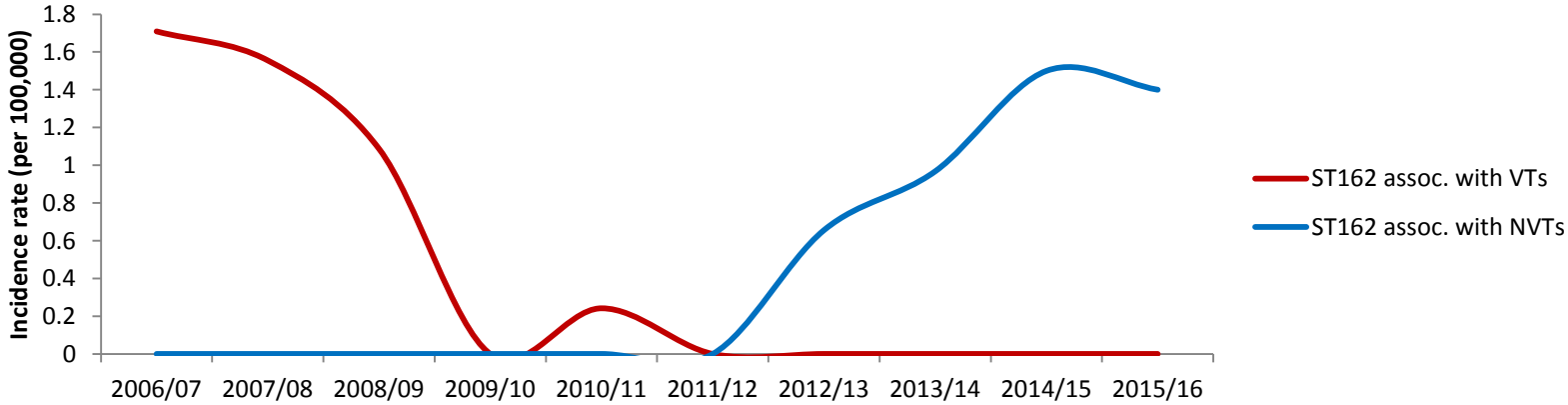


Figure 45 VT and NVT associations of IPD isolates of ST162 over time

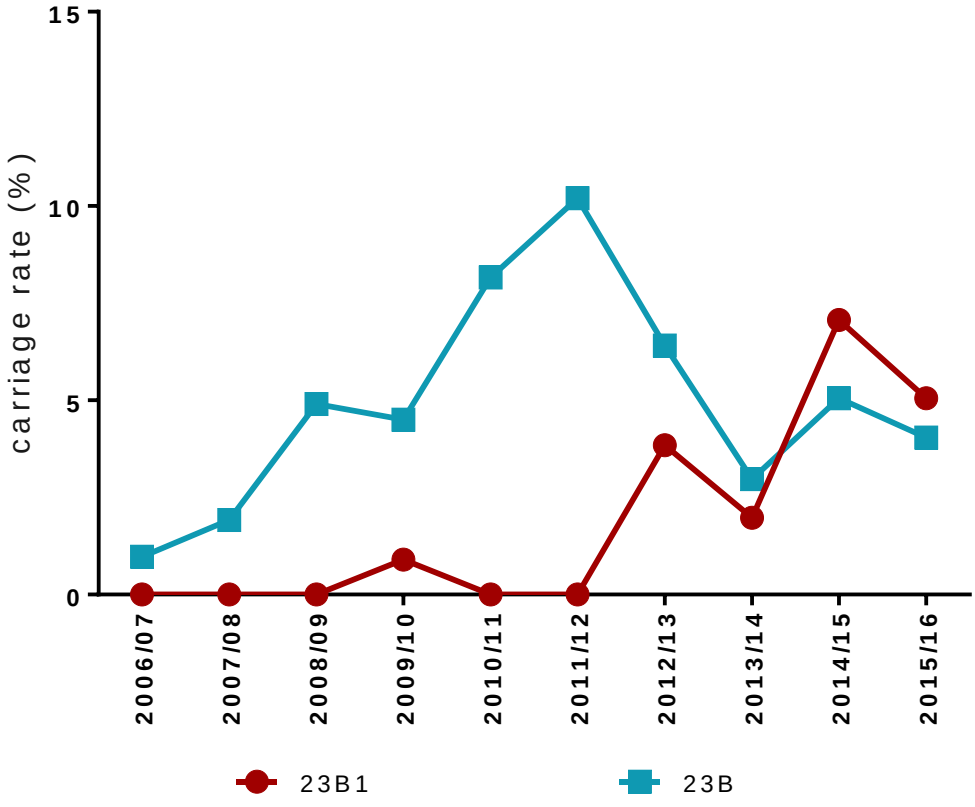


5.3.9 Molecular analysis of serotype 23B

During the course of the study a molecular subtype of serotype 23B was identified (Kapatai *et al.*, 2017b). Carriage and IPD isolates that had previously been serotyped as 23B were re-evaluated to ascertain when the subtype emerged and if its emergence was observed in both carriage and IPD.

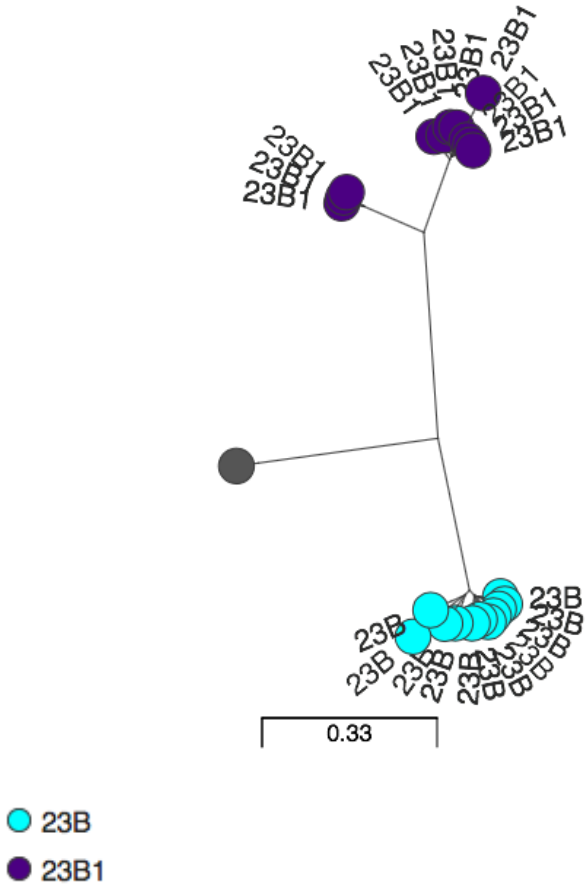
A total of 66 23B carriage isolates and 3 IPD isolates were investigated.18/66 carriage isolates were subtyped as 23B1. All of the IPD isolates were found to be 23B. Figure 46 shows the trend of 23B/23B1 carriage over time. 23B1 emerged in the carriage isolates in the fourth study year (2009/10) and has increased significantly since then (p= <0.0001).

Figure 46 Carriage of 23B/23B1 serotypes over 10 years



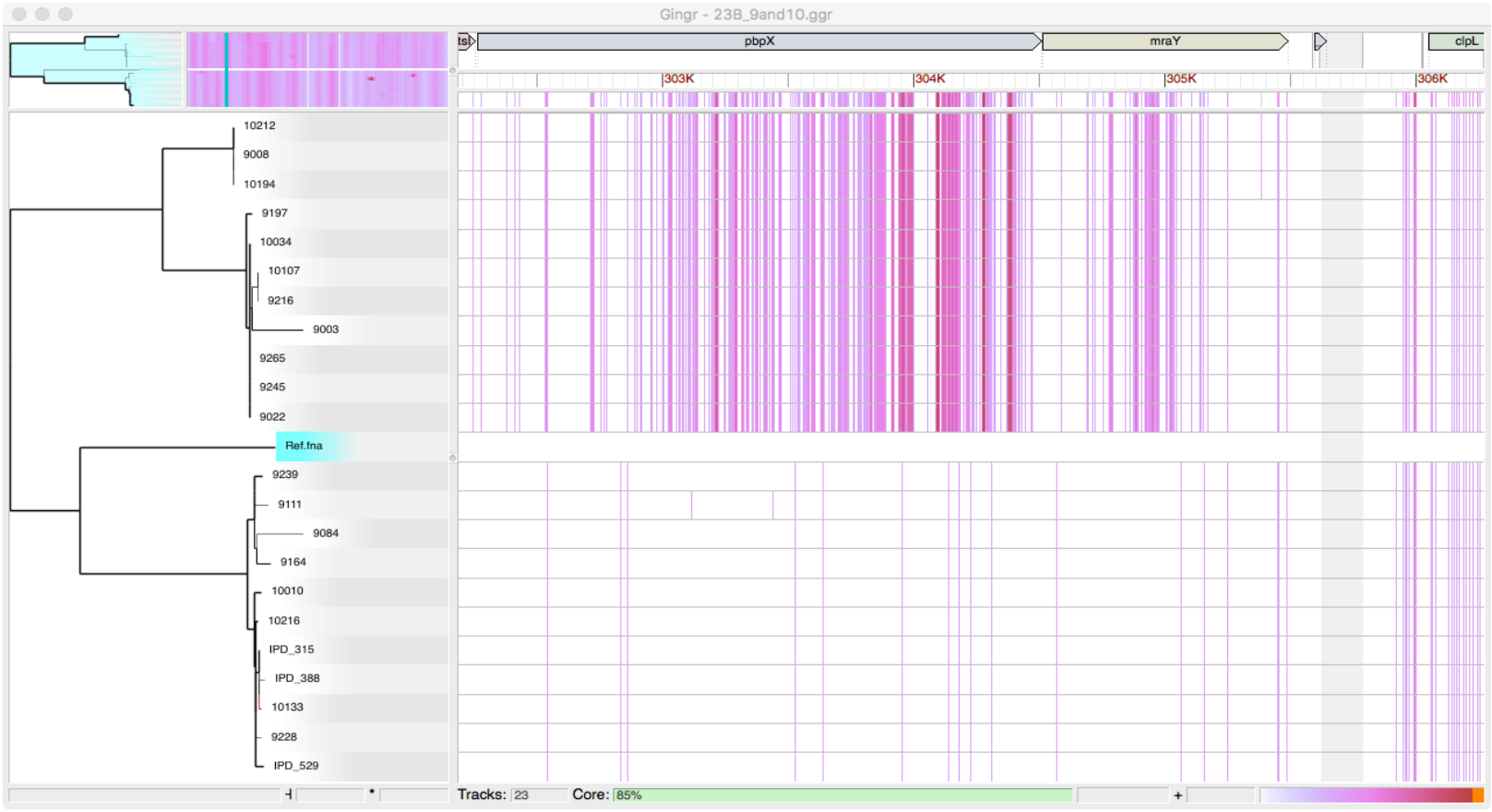
A core genome SNP tree was produced using 23B and 23B1 isolates found in carriage and IPD in years 9 and 10 (2014/15 and 2015/16). It is shown in Figure 47. The 23B1 isolates were seen to branch separately and this corresponded with different STs, 2372, 1373 and 11167.

Figure 47 Core genome SNP tree of serotype 23B/23B1 isolates



The core alignments of 23B and 23B1 isolates from study years 2014/15 and 2015/16 were interrogated for differences using Gingr (harvest citation here). A genomic region that corresponds to the pbpX gene (a penicillin binding protein) displayed variation between all 23B1 isolates and 23B isolates. Other regions of differences were observed however the pbpX gene was noted as the differences here were common to all of the subtype 23B1 isolates irrespective of ST. The output is displayed in Figure 48.

Figure 48 Genomic variation in 23B and 23B1 isolates



The upper 11 isolates were subtyped as 23B1 and the lower 11 isolates were typed as 23B. Variation between the two can be seen in the pbpX gene. The 23B1 isolates shown here include two different STs. Reference strain is ATCC700699 (serotype 23F).

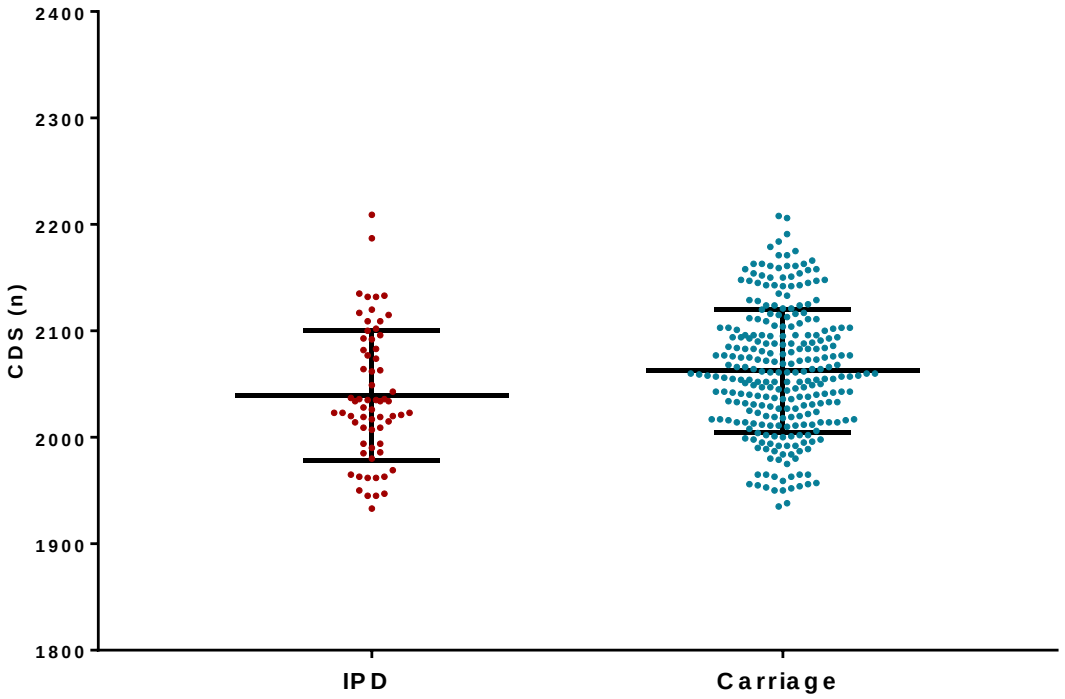
5.3.10 Genomic data

The sequencing statistics that accompanied each pneumococcal WGS included the number of coding DNA sequences (CDS) in each genome.

For this analysis only sequences from study years 8, 9 and 10 were used. The number of CDS in each IPD isolate, was compared with those in the carriage isolates. The results are shown in Figure 49.

The number of CDS per genome had a slightly lower mean value in the IPD isolates than in the carriage isolates at 2039 and 2062 respectively. Observation of serotype and number of CDS per genome found no obvious correlation.

Figure 49 Coding sequences in carriage and IPD isolates



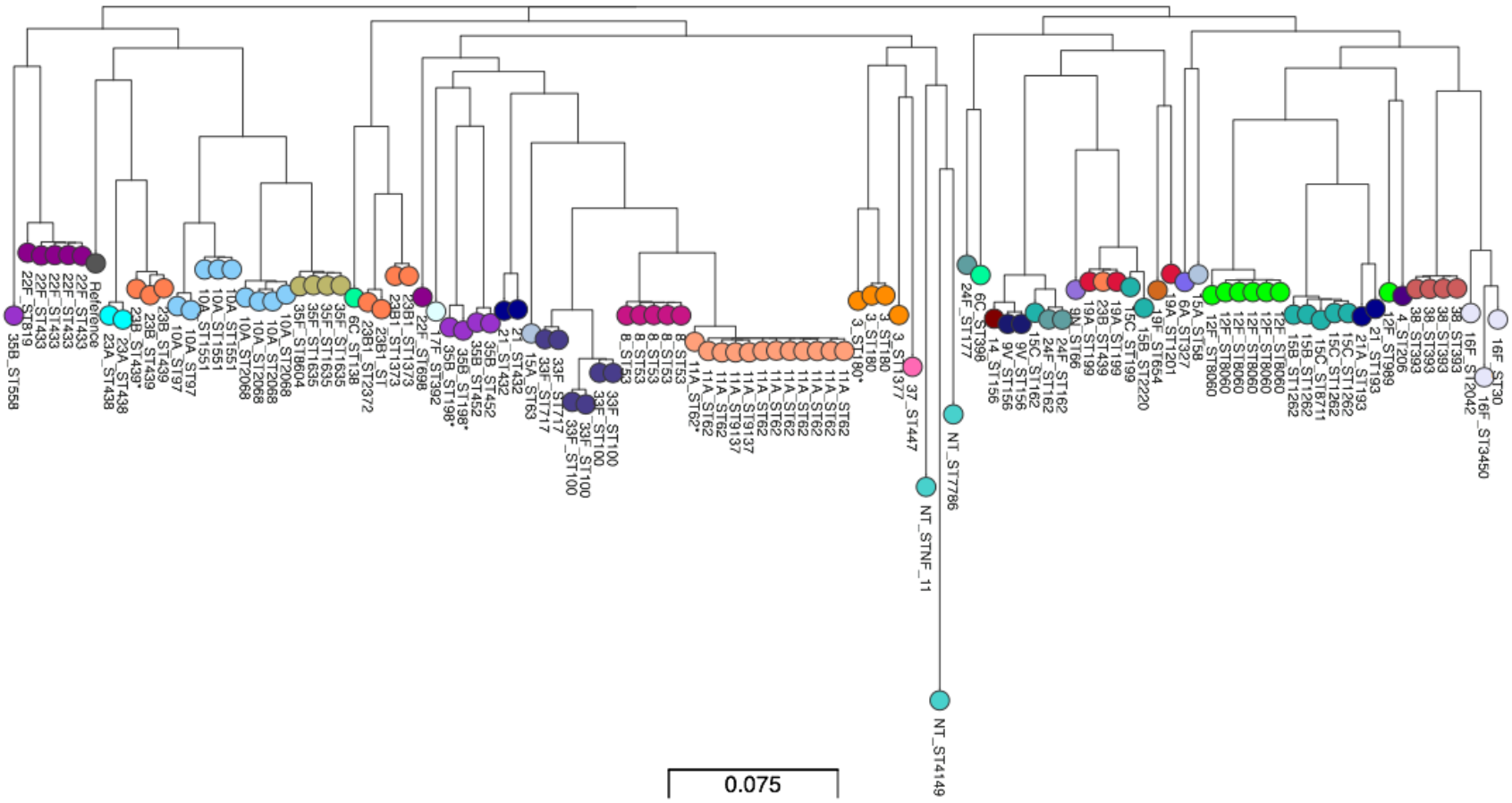
The core genomes of pneumococci isolated from carriage and IPD in 2015/16 were aligned and a phylogeny was inferred based upon SNP variations. The resultant tree is presented in Figure 50.

Isolates were observed to cluster by ST rather than serotype however some STs appeared to display earlier divergence than others. Serotype 11A was observed to have two associated STs; 62 and 9137 and both STs inhabited the same clade within the tree. Examination of MLST data found that both STs are found within the same CC and the predicted founder of ST9137 is ST62. Serotype 11A appeared in IPD five times over the 10 study years each time associated with ST62.

Serotype 35B is observed to have three associated STs; 558, 198 and 452. Whilst ST198 and ST452 are clustered within the same clade, ST558 is located within a separate clade indicating earlier divergence. ST558 and ST452 have not formed a CC with any other ST observed over the 10 years and ST198 formed a complex with a novel ST. The novel ST was observed twice in carriage in 2011/12 and serotyped as 35B both times.

Serotype 15B/C was separated into 15B and 15C after examination of a TA repeater sequence within the O-acetyltransferase gene (van Selm *et al.*, 2003a). Eight TA repeats are found in isolates designated 15B while seven or nine repeats are found in 15C isolates. This serotype was separated into two serotypes for the molecular examination in order to examine the possibility of further genomic variations. One clade which comprised entirely of 15B and 15C serotypes showed associations with two STs; ST1262 and ST8711. Examination of MLST data revealed ST1262 and ST8711 form a CC with only variation of the *gdh* allelic fragment to separate them. ST1262 appeared in carriage 29 times over the 10 years associated with 15B/C each time. It was not observed in IPD during this time. ST8711 was observed in both carriage and IPD each time associated with 15B/C.

Figure 50 Phylogeny of pneumococci isolated in 2015/16



Isolates from carriage and IPD are displayed. An * denotes a variation in one of the MLST defining genes.

5.4 Discussion

The majority of pneumococci isolated from cases of IPD had either identical STs to those found in paediatric carriage or belonged to the same clonal complexes. A number of studies have provided evidence to suggest pneumococcal carriage in children acts as a reservoir for pneumococcal clones that go on to cause IPD in the wider community (Dowling *et al.*, 1971, Walter *et al.*, 2009) however examination of molecular epidemiology shows that this connection is not straightforward. This study finds that while there are a number of similarities between carriage and IPD there are also a number of molecular variations that exist in one population but not the other.

Comparison of carriage rates and IPD incidence rates of individual STs showed that ST191 had a high incidence rate relative to other STs and a lower carriage rate. An overall decrease in ST191 was seen over the study period and this corresponds with its associated serotype, 7F which is one of the VTs included in PCV13. Evidence of the impact of PCV13 on IPD is further demonstrated by the decrease of other STs that are primarily associated with VT serotypes.

Not all STs associated with VT serotypes have decreased. Observations of ST199 over ten years have shown its continued prevalence in carriage and IPD, however within the carriage dataset the serotypes that are associated with ST199 have altered. From 2006/07 to 2011/12, ST199 was primarily associated with VT serotype 19A. From 2011/12 onwards ST199 has been associated with serotype 15B/C suggesting a possible capsular switch event at around the time of PCV13 introduction. This change has not been observed in IPD to the same extent. A single isolate of serotype 15B/C ST199 was observed pre-PCV13 (in 2008/09) however since then ST199 pneumococci have been isolated in four separate years, each time associated with VT 19A. The prevalence of NVT ST199 in carriage and VT ST199 in IPD demonstrates that the pneumococcal serotype is the prime determinant of invasive disease.

Observations of ST162 over 10 years have revealed another possible capsular switch event. ST162 was primarily associated with VT serotypes 9V and 19F in both carriage and IPD. Unlike ST199, the switch of association from VT serotypes to NVT was seen in both carriage and IPD. Serotype 24F which has increased in prevalence in carriage and disease was primarily associated with ST72 in earlier

study years however at around the time of PCV13 introduction it was more frequently found to be ST162.

The emergence of 23B1, a subtype of serotype 23B provided an opportunity to examine if the same molecular changes occurring in pneumococcal carriage were also occurring in IPD. 23B1 emerged in the fourth year of the carriage study (2009/10) and has continued to increase in numbers each year. By 2015/16 it outnumbered the 23B serotypes found in carriage. Three IPD isolates have been typed as 23B, the last one was isolated in 2015/16. To date no 23B1 isolates have been found in this IPD dataset.

The identification of SNPs within the core genome of isolates demonstrated that pneumococci generally cluster according to ST. The 23B1 isolates from 2014/15 and 2015/16 were found to be associated with three STs, 2372, 1373 and 11167. Interrogation of the core genomes of 23B and 23B1 isolates showed genetic variation common to all 23B1 isolates examined regardless of ST.

Molecular comparison of carriage and disease isolates have shown that while similarities exist regarding clonal types, a number of molecular adaptions have occurred in carriage that have not been seen in IPD. Whilst the introduction of PCV13 has altered the molecular epidemiology of both carriage and IPD the clinical implications of these changes require ongoing surveillance and further genomic investigations of both populations.

Chapter 6 Discussion

This thesis set out to test the hypothesis that population changes seen in pneumococci isolated from asymptomatic carriage during the implementation of PCV13 will also be seen in pneumococci isolated from invasive pneumococcal disease matched for location and time.

To test this hypothesis the population of pneumococci isolated from paediatric carriage was first characterised to identify changes to the carriage rate and the serotype distribution over 10 years. This was then followed by an examination of the population of pneumococci isolated from cases of IPD over the same10 years. Finally the two populations (carriage and disease) were compared to determine if underlying molecular variations could be used to identify similarities between carriage and disease and therefore infer that IPD is caused by the same population of pneumococci that is found carried asymptomatically by young children.

6.1 Key findings

Investigations involving pneumococcal carriage generally focus on children aged under 5 years. Whilst this is logical due to the higher carriage rate amongst this age group and their status as vaccine recipients, this study has shown that post-PCV13 analysis using the whole of this age group would best be performed on data collected after 2015/16. Data used prior to this date should be stratified by age to ensure only PCV13 recipients are captured in analyses that aim to measure the effect of PCV13. Knowledge of which PCV vaccine participants have received is especially important when discussing vaccine escape to ensure accuracy of data that could influence public health policies.

This study found that when analysed for trend over 10 years, pneumococcal carriage has increased. This increase has not been seen in all age groups. Carriage has remained statistically stable in infants aged 0 to 11 months and this is the group in which the impact of PCV13 can be measured over the longest time. A small increase was observed in the years that followed PCV13 introduction however from 2013/14 to 2015/16 the carriage rate has been lower. Carriage was also stable in the oldest age group 35 months to 59 months. This is the age group in which the impact of PCV13 can be measured for the least

amount of time as participants could have been recipients of PCV7 at the time that they were sampled.

There are a number of possibilities that could affect the increase in carriage seen in the middle two age groups. It is possible that the period shortly after PCV13 administration causes a temporary increase in carriage as NVT serotypes vie to fill the niche that is left when VT serotypes are targeted by a vaccine. The youngest age group and the older age group would arrive at equilibrium earlier or later than the middle age groups and the carriage data presented in this thesis may have captured the transition period. Continued surveillance will be able to provide evidence if this is the case.

The influence of participant age on the overall carriage rate should be considered alongside the study findings. Age stratification revealed differences in pneumococcal carriage between age groups. If study years were disproportionally represented by a particular age group then the overall carriage rate could be affected.

The middle age ranges, 12 to 23 months and 24 to 35 months have the highest carriage rate of the four age categories although they are only slightly higher than the oldest group. It has been shown that the administration of LAIV to the older study participants may have influenced this increase in carriage. This finding serves to highlight the importance of studying vaccine impact beyond the target of the vaccine. Other residents of the respiratory tract are affected when one member of a microbiome is altered (Bogaert *et al.*, 2011) and little is known at present on what clinical implications this might have.

Serotype distribution was found to be different in differing age categories. In 2015/16 serotypes 35B and 33F were the most prevalent in infants aged 0 to 11 months. In older children serotypes 11A and 15B/C were the most prevalent. Invasive disease has a higher incidence rate in the 0 to 2 year olds than children aged 2 to 4 years (Ladhani *et al.*, 2018) so an understanding of differences in serotype distribution in these age groups could be of importance. Comparisons with findings from other carriage studies could provide information needed to see if this is a local fluctuation or a general difference that warrants further investigation.

Analysis of local IPD cases has shown that IPD incidence has increased in recent years. This trend has been found in national IPD incidence too and reported by PHE using data derived from England and Wales which includes the data

presented within this thesis. PHE has attributed the increase in incidence to the number of cases in people aged 65 years and over caused by NVT serotypes. Analyses of national IPD data however have not found an increase in IPD in the 0 to 4 years age group as has been found in the Southampton data. Whilst IPD data local to the Southampton area lacks the statistical power of the larger national data, the recent increase in IPD in children aged 0 to 4 years must be investigated further. The increase has continued past the years of this study and may indicate a trend that will be picked up in national data later.

Compilation of national disease statistics is an important part of public health however sole reliance on these to inform public health policies could delay early recognition of important trends. Whilst large datasets have the statistical power to provide accurate assessments of disease trends their use may dilute the effects seen at local levels.

6.1.1 The impact of PCV13 on carriage and IPD

The carriage and IPD data included in this thesis covers the epidemiological years 2006/07 to 2015/16 and therefore provides information that reflects both the pre-PCV13 and the post-PCV13 eras.

Both carriage and disease caused by the additional six serotypes included in PCV13 decreased in the post-PCV13 era. The pneumococcal carriage rate for the additional VT serotypes in children aged 4 years and under was 19.8% in the year prior to PCV13 introduction (2009/10) and this decreased to 0% in 2015/16. In IPD (all ages) the incidence of disease caused by the additional VT serotypes was 4.9 cases per 100,000 in 2009/10 and this decreased to 2.7 cases per 100,000. Analysis of IPD incidence by age group showed that this decrease was primarily due to a reduction of incidence in children aged 0 to 4 years. IPD in this age group reduced from 17.5 cases per 100,000 in 2009/10 to 0 cases in 2015/16. Whilst these small numbers are difficult to statistically power, the observable decrease provides evidence that PCV13 has reduced the burden of disease caused by these serotypes in young children.

In older adults a decrease in IPD incidence caused by PCV13 VT serotype was also observed however the decrease was smaller than the one seen in young children. IPD caused by VT serotypes decreased from 9.6 cases per 100,000 in 2009/10 to 7.5 cases per 100,000 in 2015/16. Again the small numbers make statistical analysis difficult however similar trends have been reported in national data which adds credence to local findings. Ongoing surveillance of serotype-specific

IPD incidence in this age group is necessary to see if this trend is ongoing or fluctuation.

An increase in carriage and disease incidence caused by NVT serotypes was observed. A serotype-specific comparison of the carriage population and the IPD population was undertaken to investigate the possibility that the population of pneumococci carried by young children acts as a reservoir for isolates that cause disease in the wider community.

A total of 22 NVT serotypes were observed to occur in carriage and disease however a comparison of serotype frequency in carriage with frequency in disease found that 11 of these serotypes had statistically significant odds of featuring in one population over the other. Three NVT serotypes had greater odds of causing IPD than carriage: 12F, 8 and 9N. A recent PHE publication has reported an increase in IPD caused by the same three serotypes (Ladhani *et al.*, 2018) which indicates that this is not just a local finding. Carriage of 12F, 8 and 9N has occurred during the 10 study years however in few years and low numbers. Community-wide carriage studies could provide valuable information regarding the transmission of these serotypes.

A study into the molecular epidemiology of pneumococci was performed in order to further investigate the relatedness of carriage and disease isolates. The current number of MLST profiles held on the pubMLST database is 13,766 (as of 27/02/2018). A total of 177 STs were observed amongst the carriage isolates and 82 STs were seen in the IPD isolates. A comparison of ST distribution in both populations (carriage and disease) revealed that 51 STs appeared in both. A total of 83% of IPD isolates had a ST that was also found in carriage. A comparison with ST-specific data from other regions is necessary to understand if this finding is evidence of possible transmission between carriage and disease or whether these STs are common to many regional populations.

During the 10 years of this study a number of molecular changes emerged amongst some subsets of pneumococci. This provided an opportunity to investigate if the same changes were occurring in carriage and invasive disease. Two possible capsular switch events were found to occur within the carriage population. ST199 and ST162 which were primarily associated with VT serotypes in the early carriage period switched to NVT associations in the later study period. Whilst the same change was seen in the IPD population for ST162 isolates, ST199 isolates did not show the same switch and continued to be associated with VT serotypes throughout the 10 years.

The emergence of 23B1 as a subtype of serotype 23B provided a further opportunity to compare molecular changes in carriage and disease. 23B1 rapidly became more prevalent in carriage than 23B however only 23B isolates have been identified in IPD. The small number of 23B pneumococci that have been isolated from local IPD cases make conclusions difficult and continued surveillance is necessary to see what the clinical implications of these molecular changes are.

6.1.2 Limitations of the study

It is possible that pneumococcal carriage has been under represented in this study. Microbiological techniques were used to isolate and identify pneumococci however the use of PCR or microarray may have yielded not only greater numbers of pneumococci but could have also shown whether multiple serotypes had colonised individual study participants.

For the calculations of IPD incidence, population data from the 2011 census was used. Estimates of annual changes to population were not known and therefore could not be accounted for. Incidence rates were calculated per 100,000 despite the population of individual age groups being below this number. This scale was chosen for ease of comparison with PHE IPD data.

Calculations involving serotype specific IPD incidence had to be adjusted to allow for missing data. The method chosen was utilised by PHE and based upon assumptions that missing data matched known data. All adjustments were made by age group and by year to ensure results were as accurate as possible. Previous studies published by PHE have used this method when only 48% of data were available. As this study had available data for 71.5% of isolates it was deemed an acceptable method.

6.2 Future work

Collection of carriage and IPD isolates has continued past the timeframe of the study presented within this thesis. Trends that have been observed for the 10 years between 2006/07 and 2015/16 can be tested through the examination of data collected for the epidemiological year 2016/17.

Over the course of the study a huge amount of genomic data has been collected. WGS for carriage and IPD isolates have already been generated and detailed genomic analyses could yield important findings. Through comparisons of genetic variations isolates can be examined for evolutionary paths, differences in

virulence and antibiotic resistance. The ever increasing number of bioinformatic tools that are publically available mean that such analyses could be performed rapidly. Having isolates from carriage and disease that are matched for time and location makes this type of analysis a logical next step.

A number of questions have arisen over the course of the study that would require a change in methods to investigate. The possibility that carriage has been under represented could be addressed with the use of PCR or microarray to identify pneumococci. This would be particularly useful to identify multiple serotypes within one host.

The issue of antibiotic resistance (AMR) is of ongoing concern to public health. Carriage isolates collected since 2006/07 have recently been tested against a panel of antibiotics used to treat pneumococcal infections. This data must now be examined and resistant organisms identified for further investigation. Changes in resistance can be correlated with the phenotypic and genotypic data to provide a comprehensive report on AMR within the carriage population.

Differences in serotype distribution between carriage and IPD and between hosts of different ages can be investigated further by examining data derived from community wide carriage studies and from analysis of pneumococci derived from cases of non-invasive pneumococcal disease. Little is known about pneumococci in the wider community and examination of serotype distribution within different cohorts may provide information relating to the transmission of serotypes identified as having increased invasive potential.

Finally, the carriage study has provided an opportunity to examine other bacterial species of interest. Isolates of *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Staphylococcus aureus* have been collected alongside pneumococcal isolates and provide an opportunity to examine the effect of PCV13 on other residents of the respiratory niche.

6.3 Concluding remarks

This study examined the impact of PCV13 on pneumococcal carriage and disease. The data within this thesis has shown that the introduction of PCV13 has been successful in terms of reducing carriage and disease caused by PCV13 VT serotypes however the increase in NVT serotypes in both carriage and disease

have almost negated this effect. IPD incidence in the Southampton area was higher in 2015/16 than in the year prior to PCV13 introduction. This increase has been driven by an increase in NVT IPD in adults aged 65 years and over.

IPD in children 0 to 4 years was lower in 2015/16 than in the year prior to PCV13 introduction however incidence in this age group has increased since 2013/14. The small number of isolates has made this analysis difficult to power and continued surveillance is needed to ascertain the accuracy of this trend.

Appendix A

List of publications that has used data derived from the carriage study:

1. Gladstone, R.A., *et al.*, Pre-vaccine serotype composition within a lineage signposts its serotype replacement – a carriage study over 7 years following pneumococcal conjugate vaccine use in the UK. Microbial Genomics, 2017. 3(6).

2. Devine, V.T., *et al.*, The rise and fall of pneumococcal serotypes carried in the PCV era. Vaccine, 2017. 35(9): p. 1293-1298.

3. Gladstone, R.A., *et al.*, Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. Vaccine, 2015. 33(17): p. 2015-21.

4. Coughtrie, A.L., *et al.*, Evaluation of swabbing methods for estimating the prevalence of bacterial carriage in the upper respiratory tract: a cross sectional study. BMJ Open, 2014. 4(10): p. e005341.

5. Loman, N.J., *et al.*, Clonal expansion within pneumococcal serotype 6C after use of seven-valent vaccine. PLoS One, 2013. 8(5): p. e64731.

6. Tocheva, A.S., *et al.*, Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of Streptococcus pneumoniae in young children. Vaccine, 2011. 29(26): p. 4400-4.

7. Tocheva, A.S., *et al.*, Distribution of carried pneumococcal clones in UK children following the introduction of the 7-valent pneumococcal conjugate vaccine: a 3-year cross-sectional population based analysis. Vaccine, 2013. 31(31): p. 3187-90.

8. Tocheva, A.S., *et al.*, Increase in serotype 6C pneumococcal carriage, United Kingdom. Emerg Infect Dis, 2010. 16(1): p. 154-5.

List of References

ADDERSON, E. E. 2001. Antibody repertoires in infants and adults: effects of T-independent and T-dependent immunizations. *Springer Semin Immunopathol,* 23**,** 387-403.

ANDAM, C. P., MITCHELL, P. K., CALLENDRELLO, A., CHANG, Q., CORANDER, J., CHAGUZA, C., MCGEE, L., BEALL, B. W. & HANAGE, W. P. 2017. Genomic Epidemiology of Penicillin-Nonsusceptible Pneumococci with Nonvaccine Serotypes Causing Invasive Disease in the United States. *Journal of Clinical Microbiology,* 55**,** 1104-1115.

ANDERSON, S. 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res,* 9**,** 3015-27.

ANSORGE, W., SPROAT, B. S., STEGEMANN, J. & SCHWAGER, C. 1986. A non-radioactive automated method for DNA sequence determination. *J Biochem Biophys Methods,* 13**,** 315-23.

APPELBAUM, P. C. 2002. Resistance among Streptococcus pneumoniae: Implications for Drug Selection. *Clinical Infectious Diseases,* 34**,** 1613-1620.

ARGIMÓN, S., ABUDAHAB, K., GOATER, R. J., FEDOSEJEV, A., BHAI, J., GLASNER, C., FEIL, E. J., HOLDEN, M. T., YEATS, C. A. & GRUNDMANN, H. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microbial Genomics,* 2.

AUSTRIAN, R. 1975. Random gleanings from a life with the pneumococcus. *The Journal of infectious diseases,* 131**,** 474-484.

AUSTRIAN, R. 1980. *Surveillance of pneumococcal infection for field trials of polyvalent pneumococcal vaccines*, National Technical Information Service.

AUSTRIAN, R. 1981. Pneumococcus: the first one hundred years. *Rev Infect Dis,* 3**,** 183-9.

AUSTRIAN, R. 1999. A Brief History of Pneumococcal Vaccines. *Drugs & Aging,* 15**,** 1-10.

AUSTRIAN, R. & GOLD, J. 1964. PNEUMOCOCCAL BACTEREMIA WITH ESPECIAL REFERENCE TO BACTEREMIC PNEUMOCOCCAL PNEUMONIA. *Ann Intern Med,* 60**,** 759-76.

AVERY, O. T., CHICKERING, H., COLE, R. & DOCHEZ, A. 1918. Acute lobar pneumonia. *THE ROCKEFELLER INSTITUTE***,** 483.

AVERY, O. T., MACLEOD, C. M. & MCCARTY, M. 1944. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES: INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *The Journal of Experimental Medicine,* 79**,** 137-158.

AVERY, O. T. & MORGAN, H. J. 1924. THE OCCURRENCE OF PEROXIDE IN CULTURES OF PNEUMOCOCCUS. *The Journal of Experimental Medicine,* 39**,** 275-287.

BARCUS, V. A., GHANEKAR, K., YEO, M., COFFEY, T. J. & DOWSON, C. G. 1995. Genetics of high level penicillin resistance in clinical isolates of Streptococcus pneumoniae. *FEMS microbiology letters,* 126**,** 299-303.

BARKER, J. H., MUSHER, D. M., SILBERMAN, R., PHAN, H. M. & WATSON, D. A. 1999. Genetic relatedness among nontypeable pneumococci implicated in sporadic cases of conjunctivitis. *Journal of clinical microbiology,* 37**,** 4039-4041.

BAROCCHI, M. A., RIES, J., ZOGAJ, X., HEMSLEY, C., ALBIGER, B., KANTH, A., DAHLBERG, S., FERNEBRO, J., MOSCHIONI, M., MASIGNANI, V., HULTENBY, K., TADDEI, A. R., BEITER, K., WARTHA, F., VON EULER, A., COVACCI, A., HOLDEN, D. W., NORMARK, S., RAPPUOLI, R. & HENRIQUES-NORMARK, B. 2006. A pneumococcal pilus influences virulence and host inflammatory responses. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 2857-2862.

BAUMDICKER, F., HESS, W. R. & PFAFFELHUBER, P. 2012. The Infinitely Many Genes Model for the Distributed Genome of Bacteria. *Genome Biology and Evolution,* 4**,** 443-456.

BENTLEY, S. D., AANENSEN, D. M., MAVROIDI, A., SAUNDERS, D., RABBINOWITSCH, E., COLLINS, M., DONOHOE, K., HARRIS, D., MURPHY, L., QUAIL, M. A., SAMUEL, G., SKOVSTED, I. C., KALTOFT, M. S., BARRELL, B., REEVES, P. R., PARKHILL, J. & SPRATT, B. G. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet,* 2**,** e31.

BERGMANN, S. & HAMMERSCHMIDT, S. 2006. Versatility of pneumococcal surface proteins. *Microbiology,* 152**,** 295-303.

BERNHEIMER, H. P. & TIRABY, J.-G. 1976. Inhibition of phage infection by pneumococcus capsule. *Virology,* 73**,** 308-309.

BEZANÇON, F. & GRIFFON, V. 1897. Pouvoir agglutinatif du sérum dans les infections expérimentales et humaines à pneumocoques. . *I C R Soc Biol,* 49**,** 551.

BLOMBERG, C., DAGERHAMN, J., DAHLBERG, S., BROWALL, S., FERNEBRO, J., ALBIGER, B., MORFELDT, E., NORMARK, S. & HENRIQUES-NORMARK, B. 2009. Pattern of accessory regions and invasive disease potential in Streptococcus pneumoniae. *The Journal of infectious diseases,* 199**,** 1032-1042.

BLUESTONE, C. D., STEPHENSON, J. S. & MARTIN, L. M. 1992. Ten-year review of otitis media pathogens. *The Pediatric infectious disease journal,* 11**,** S7-11.

BOGAERT, D., DE GROOT, R. & HERMANS, P. W. 2004a. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis,* 4**,** 144-54.

BOGAERT, D., DE GROOT, R. & HERMANS, P. W. M. 2004b. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *The Lancet Infectious Diseases,* 4**,** 144-154.

BOGAERT, D., ENGELEN, M. N., TIMMERS-REKER, A. J., ELZENAAR, K. P., PEERBOOMS, P. G., COUTINHO, R. A., DE GROOT, R. & HERMANS, P. W.

2001. Pneumococcal carriage in children in The Netherlands: a molecular epidemiological study. *Journal of clinical microbiology,* 39**,** 3316-3320.

BOGAERT, D., KEIJSER, B., HUSE, S., ROSSEN, J., VEENHOVEN, R., VAN GILS, E., BRUIN, J., MONTIJN, R., BONTEN, M. & SANDERS, E. 2011. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PloS one,* 6**,** e17035.

BOGAERT, D., VAN BELKUM, A., SLUIJTER, M., LUIJENDIJK, A., DE GROOT, R., RÜMKE, H. C., VERBRUGH, H. A. & HERMANS, P. W. M. 2004c. Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children. *The Lancet,* 363**,** 1871-1872.

BOGAERT, D., WEINBERGER, D., THOMPSON, C., LIPSITCH, M. & MALLEY, R. 2009. Impaired Innate and Adaptive Immunity to Streptococcus pneumoniae and Its Effect on Colonization in an Infant Mouse Model. *Infection and Immunity,* 77**,** 1613-1622.

BRATCHER, P. E., KIM, K.-H., KANG, J. H., HONG, J. Y. & NAHM, M. H. 2010. Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization. *Microbiology,* 156**,** 555-560.

BRUEGGEMANN, A. B., GRIFFITHS, D. T., MEATS, E., PETO, T., CROOK, D. W. & SPRATT, B. G. 2003. Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis,* 187**,** 1424-32.

BRUEGGEMANN, A. B., PAI, R., CROOK, D. W. & BEALL, B. 2007. Vaccine Escape Recombinants Emerge after Pneumococcal Vaccination in the United States. *PLoS Pathogens,* 3**,** e168.

BRYANT, J. M., GROGONO, D. M., GREAVES, D., FOWERAKER, J., RODDICK, I., INNS, T., REACHER, M., HAWORTH, C. S., CURRAN, M. D., HARRIS, S. R., PEACOCK, S. J., PARKHILL, J. & FLOTO, R. A. 2013. Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study. *The Lancet,* 381**,** 1551-1560.

BUCHANAN, R. & GIBBONS, N. E. 1974. Bergey's manual of determinative bacteriology. *8th edition. Baltimore: Williams & Wilkins***,** 490-517.

BUMGARNER, R. 2013. DNA microarrays: Types, Applications and their future. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.],* 0 22**,** Unit-22.1.

BURRUS, V., PAVLOVIC, G., DECARIS, B. & GUÉDON, G. 2002. Conjugative transposons: the tip of the iceberg. *Molecular Microbiology,* 46**,** 601-610.

BUTLER, J. C., SHAPIRO, E. D. & CARLONE, G. M. 1999. Pneumococcal vaccines: history, current status, and future directions. *The American Journal of Medicine,* 107**,** 69-76.

BYINGTON, C. L., SPENCER, L. Y., JOHNSON, T. A., PAVIA, A. T., ALLEN, D., MASON, E. O., KAPLAN, S., CARROLL, K. C., DALY, J. A. & CHRISTENSON, J. C. 2002. An epidemiological investigation of a sustained high rate of

pediatric parapneumonic empyema: risk factors and microbiological associations. *Clinical Infectious Diseases*, 434-440.

CALIX, J. J. & NAHM, M. H. 2010. A new pneumococcal serotype, 11E, has a variably inactivated wcjE gene. *J Infect Dis,* 202, 29-38.

CALIX, J. J., PORAMBO, R. J., BRADY, A. M., LARSON, T. R., YOTHER, J., ABEYGUNWARDANA, C. & NAHM, M. H. 2012. Biochemical, Genetic, and Serological Characterization of Two Capsule Subtypes among Streptococcus pneumoniae Serotype 20 Strains DISCOVERY OF A NEW PNEUMOCOCCAL SEROTYPE. *Journal of Biological Chemistry,* 287, 27885-27894.

CARTEE, R. T., FORSEE, W. T., SCHUTZBACH, J. S. & YOTHER, J. 2000. Mechanism of Type 3 Capsular Polysaccharide Synthesis inStreptococcus pneumoniae. *Journal of Biological Chemistry,* 275, 3907-3914.

CARVALHO, M. D. G., PIMENTA, F. C., MOURA, I., ROUNDTREE, A., GERTZ, R. E., JR., LI, Z., JAGERO, G., BIGOGO, G., JUNGHAE, M., CONKLIN, L., FEIKIN, D. R., BREIMAN, R. F., WHITNEY, C. G. & BEALL, B. W. 2013. Non-pneumococcal mitis-group streptococci confound detection of pneumococcal capsular serotype-specific loci in upper respiratory tract. *PeerJ,* 1, e97.

CASTRO-WALLACE, S. L., CHIU, C. Y., JOHN, K. K., STAHL, S. E., RUBINS, K. H., MCINTYRE, A. B., DWORKIN, J. P., LUPISELLA, M. L., SMITH, D. J. & BOTKIN, D. J. 2016. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *bioRxiv*, 077651.

CHÂTELET, I. P. D., TRAORE, Y., GESSNER, B. D., ANTIGNAC, A., NACCRO, B., NJANPOP-LAFOURCADE, B.-M., OUEDRAOGO, M. S., TIENDREBEOGO, S. R., VARON, E. & TAHA, M. K. 2005. Bacterial Meningitis in Burkina Faso: Surveillance Using Field-Based Polymerase Chain Reaction Testing. *Clinical Infectious Diseases,* 40, 17-25.

CHEWAPREECHA, C., HARRIS, S. R., CROUCHER, N. J., TURNER, C., MARTTINEN, P., CHENG, L., PESSIA, A., AANENSEN, D. M., MATHER, A. E., PAGE, A. J., SALTER, S. J., HARRIS, D., NOSTEN, F., GOLDBLATT, D., CORANDER, J., PARKHILL, J., TURNER, P. & BENTLEY, S. D. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet,* 46, 305-309.

CHIDGEAVADZE, Z. G., BEABEALASHVILLI, R. S., ATRAZHEV, A. M., KUKHANOVA, M. K., AZHAYEV, A. V. & KRAYEVSKY, A. A. 1984. 2',3'-Dideoxy-3' aminonucleoside 5'-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases. *Nucleic Acids Res,* 12, 1671-86.

CLAVERYS, J.-P., GROSSIORD, B. & ALLOING, G. 2000. Is the Ami-AliA/B oligopeptide permease of Streptococcus pneumoniae involved in sensing environmental conditions? *Research in microbiology,* 151, 457-463.

CLAVERYS, J.-P., PRUDHOMME, M. & MARTIN, B. 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annual review of microbiology,* 60.

CLAVERYS, J. P., MARTIN, B. & HÅVARSTEIN, L. S. 2007. Competence-induced fratricide in streptococci. *Molecular microbiology,* 64, 1423-1433.

CLEARY, D. W., DEVINE, V. T., JEFFERIES, J. M. C., WEBB, J. S., BENTLEY, S. D., GLADSTONE, R. A., FAUST, S. N. & CLARKE, S. C. 2016. Comparative Genomics of Carriage and Disease Isolates of Streptococcus pneumoniae Serotype 22F Reveals Lineage-Specific Divergence and Niche Adaptation. *Genome Biology and Evolution,* 8**,** 1243-1251.

COLLINS, R. E. & HIGGS, P. G. 2012. Testing the Infinitely Many Genes Model for the Evolution of the Bacterial Core Genome and Pangenome. *Molecular Biology and Evolution,* 29**,** 3413-3425.

COOPER, G., ROSENSTEIN, C., WALTER, A. & PEIZER, L. 1932. THE FURTHER SEPARATION OF TYPES AMONG THE PNEUMOCOCCI HITHERTO INCLUDED IN GROUP IV AND THE DEVELOPMENT OF THERAPEUTIC ANTISERA FOR THESE TYPES. *J Exp Med,* 55**,** 531-54.

CORANDER, J., FRASER, C., GUTMANN, M. U., ARNOLD, B., HANAGE, W. P., BENTLEY, S. D., LIPSITCH, M. & CROUCHER, N. J. 2017. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nature Ecology & Evolution,* 1**,** 1950-1960.

CRIM, C., CALVERLEY, P., ANDERSON, J., CELLI, B., FERGUSON, G., JENKINS, C., JONES, P., WILLITS, L., YATES, J. & VESTBO, J. 2009. Pneumonia risk in COPD patients receiving inhaled corticosteroids alone or in combination: TORCH study results. *European Respiratory Journal,* 34**,** 641-647.

CROUCHER, N. J., COUPLAND, P. G., STEVENSON, A. E., CALLENDRELLO, A., BENTLEY, S. D. & HANAGE, W. P. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nature Communications,* 5**,** 5471.

CROUCHER, N. J., HARRIS, S. R., FRASER, C., QUAIL, M. A., BURTON, J., VAN DER LINDEN, M., MCGEE, L., VON GOTTBERG, A., SONG, J. H., KO, K. S., PICHON, B., BAKER, S., PARRY, C. M., LAMBERTSEN, L. M., SHAHINAS, D., PILLAI, D. R., MITCHELL, T. J., DOUGAN, G., TOMASZ, A., KLUGMAN, K. P., PARKHILL, J., HANAGE, W. P. & BENTLEY, S. D. 2011a. Rapid pneumococcal evolution in response to clinical interventions. *Science,* 331**,** 430-4.

CROUCHER, N. J., VERNIKOS, G. S., PARKHILL, J. & BENTLEY, S. D. 2011b. Identification, variation and transcription of pneumococcal repeat sequences. *BMC Genomics,* 12**,** 120.

CROUCHER, N. J., WALKER, D., ROMERO, P., LENNARD, N., PATERSON, G. K., BASON, N. C., MITCHELL, A. M., QUAIL, M. A., ANDREW, P. W., PARKHILL, J., BENTLEY, S. D. & MITCHELL, T. J. 2009a. Role of Conjugative Elements in the Evolution of the Multidrug-Resistant Pandemic Clone Streptococcus pneumoniae(Spain23F) ST81. *Journal of Bacteriology,* 191**,** 1480-1489.

CROUCHER, N. J., WALKER, D., ROMERO, P., LENNARD, N., PATERSON, G. K., BASON, N. C., MITCHELL, A. M., QUAIL, M. A., ANDREW, P. W., PARKHILL, J., BENTLEY, S. D. & MITCHELL, T. J. 2009b. Role of Conjugative Elements in the Evolution of the Multidrug-Resistant Pandemic Clone Streptococcus pneumoniaeSpain23F ST81. *Journal of Bacteriology,* 191**,** 1480-1489.

CUNDELL, D. R., GERARD, N. P., GERARD, C., IDANPAAN-HEIKKILA, I. & TUOMANEN, E. I. 1995. Streptococcus pneumoniae anchor to activated human cells by the receptor for platelet-activating factor. *Nature,* 377**,** 435-438.

DAVIS, S. M., DELORIA-KNOLL, M., KASSA, H. T. & O'BRIEN, K. L. 2013. Impact of pneumococcal conjugate vaccines on nasopharyngeal carriage and invasive disease among unvaccinated people: Review of evidence on indirect effects. *Vaccine,* 32**,** 133-145.

DELIHAS, N. 2008. Small mobile sequences in bacteria display diverse structure/function motifs. *Molecular Microbiology,* 67**,** 475-481.

DELVES , P. J. & ROITT , I. M. 2000. The Immune System. *New England Journal of Medicine,* 343**,** 37-49.

DEVINE, V. T., CLEARY, D. W., JEFFERIES, J. M. C., ANDERSON, R., MORRIS, D. E., TUCK, A. C., GLADSTONE, R. A., O'DOHERTY, G., KURUPARAN, P., BENTLEY, S. D., FAUST, S. N. & CLARKE, S. C. 2017. The rise and fall of pneumococcal serotypes carried in the PCV era. *Vaccine,* 35**,** 1293-1298.

DILLARD, J. P., VANDERSEA, M. W. & YOTHER, J. 1995. Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in Streptococcus pneumoniae. *J Exp Med,* 181**,** 973-83.

DONATI, C., HILLER, N. L., TETTELIN, H., MUZZI, A., CROUCHER, N. J., ANGIUOLI, S. V., OGGIONI, M., HOTOPP, J. C. D., HU, F. Z. & RILEY, D. R. 2010. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome biology,* 11**,** R107.

DOPAZO, J., MENDOZA, A., HERRERO, J., CALDARA, F., HUMBERT, Y., FRIEDLI, L., GUERRIER, M., GRAND-SCHENK, E., GANDIN, C. & DE FRANCESCO, M. 2001. Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate. *Microbial drug resistance,* 7**,** 99-125.

DOWLING, J. N., SHEEHE, P. R. & FELDMAN, H. A. 1971. Pharyngeal Pneumococcal Acquisitions in "Normal" Families: A Longitudinal Study. *The Journal of Infectious Diseases,* 124**,** 9-17.

DOWSON, C. G., COFFEY, T. J., KELL, C. & WHILEY, R. A. 1993. Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae. *Molecular microbiology,* 9**,** 635-643.

DOWSON, C. G., COFFEY, T. J. & SPRATT, B. G. 1994. Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to β-lactam antibiotics. *Trends in microbiology,* 2**,** 361-366.

DUNAIS, B., PRADIER, C., CARSENTI, H., SABAH, M., MANCINI, G., FONTAS, E. & DELLAMONICA, P. 2003. Influence of child care on nasopharyngeal carriage of Streptococcus pneumoniae and Haemophilus influenzae. *The Pediatric infectious disease journal,* 22**,** 589-593.

ELBERSE, K., WITTEVEEN, S., VAN DER HEIDE, H., VAN DE POL, I., SCHOT, C., VAN DER ENDE, A., BERBERS, G. & SCHOULS, L. 2011. Sequence diversity within the capsular genes of Streptococcus pneumoniae serogroup 6 and 19. *PLoS One,* 6**,** e25018.

ENRIGHT, M. C. & SPRATT, B. G. 1998. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology,* 144**,** 3049-3060.

EVANS, G. & GAISFORD, W. F. 1938. Treatment of Pneumonia with 2-(p-Aminobenzenesul-phonamido) Pyridine. *Lancet*, 14-19.

EYRE, D. W., CULE, M. L., WILSON, D. J., GRIFFITHS, D., VAUGHAN, A., O'CONNOR, L., IP, C. L., GOLUBCHIK, T., BATTY, E. M. & FINNEY, J. M. 2013. Diverse sources of C. difficile infection identified on whole-genome sequencing. *New England Journal of Medicine,* 369, 1195-1205.

EYRE, J. W. H. & WASHBOURN, J. W. 1899. Further experiments with Pane's anti-pneumococcus serum. . *Br Med J* 2, 1247-1249.

FEIKIN, D. R., JAGERO, G., AURA, B., BIGOGO, G. M., OUNDO, J., BEALL, B. W., KARANI, A., MORPETH, S., NJENGA, M. K. & BREIMAN, R. F. 2010. High rate of pneumococcal bacteremia in a prospective cohort of older children and adults in an area of high HIV prevalence in rural western Kenya. *BMC Infect Dis,* 10, 186.

FEIL, E. J., LI, B. C., AANENSEN, D. M., HANAGE, W. P. & SPRATT, B. G. 2004. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *Journal of Bacteriology,* 186, 1518-1530.

FELDMAN, C. & ANDERSON, R. 2014. Recent advances in our understanding of Streptococcus pneumoniae infection. *F1000Prime Rep,* 6, 82.

FINLAND, M., GARNER, C., WILCOX, C. & SABATH, L. D. 1976. Susceptibility of pneumococci and Haemophilus influenzae to antibacterial agents. *Antimicrobial agents and chemotherapy,* 9, 274-287.

FLASCHE, S., VAN HOEK, A. J., SHEASBY, E., WAIGHT, P., ANDREWS, N., SHEPPARD, C., GEORGE, R. & MILLER, E. 2011. Effect of Pneumococcal Conjugate Vaccination on Serotype-Specific Carriage and Invasive Disease in England: A Cross-Sectional Study. *PLOS Medicine,* 8, e1001017.

FRAENKEL, A. 1884. Die genuine Pneumonie. *Verh Cong Inn Med,* 3, 17-31.

FRAENKEL, A. 1886. Weitere Beitrage zur Lehre von den Mikrococcen der genuinen fibrinosen Pneumonie. *Zeitschrift filr Klinische Medicin 1886b*, 437-58.

FRANCISCO, A. P., VAZ, C., MONTEIRO, P. T., MELO-CRISTINO, J., RAMIREZ, M. & CARRIÇO, J. A. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics,* 13, 87.

FRIEDLANDER, C. 1883. Die Mikrokokken der Pneumonie. *Fortschritte der Medizin (Munchen),* 58, 475-81.

GARAU, J., LINARES, J. & DOMINGUEZ, C. 1981. Chloramphenicol-resistant pneumococci. *The Lancet,* 318, 147-148.

GASPAR, P., AL-BAYATI, F. A. Y., ANDREW, P. W., NEVES, A. R. & YESILKAYA, H. 2014. Lactate Dehydrogenase Is the Key Enzyme for Pneumococcal Pyruvate Metabolism and Pneumococcal Survival in Blood. *Infection and Immunity,* 82, 5099-5109.

GBD 2016. Institute of Health Metrics and Evaluation (IHME), GBD Results Tool, Seatle WA. accessed 22/11/17.

GENO, K. A., GILBERT, G. L., SONG, J. Y., SKOVSTED, I. C., KLUGMAN, K. P., JONES, C., KONRADSEN, H. B. & NAHM, M. H. 2015. Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clinical Microbiology Reviews,* 28**,** 871-899.

GENO, K. A., SAAD, J. S. & NAHM, M. H. 2017. Discovery of Novel Pneumococcal Serotype 35D, a Natural WciG-Deficient Variant of Serotype 35B. *Journal of Clinical Microbiology,* 55**,** 1416-1425.

GERTZ JR, R., LI, Z., PIMENTA, F., JACKSON, D., JUNI, B., LYNFIELD, R., JORGENSEN, J. & CARVALHO MDAG, B. B. 2010. Active Bacterial Core Surveillance Team. Increased penicillin nonsusceptibility of nonvaccine-serotype invasive pneumococci other than serotypes 19A and 6A in post-7-valent conjugate vaccine era. *J Infect Dis,* 201**,** 770-775.

GLADSTONE, R. A., DEVINE, V., JONES, J., CLEARY, D., JEFFERIES, J. M., BENTLEY, S. D., FAUST, S. N. & CLARKE, S. C. 2017. Pre-vaccine serotype composition within a lineage signposts its serotype replacement – a carriage study over 7 years following pneumococcal conjugate vaccine use in the UK. *Microbial Genomics,* 3.

GLADSTONE, R. A., JEFFERIES, J. M., TOCHEVA, A. S., BEARD, K. R., GARLEY, D., CHONG, W. W., BENTLEY, S. D., FAUST, S. N. & CLARKE, S. C. 2015. Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. *Vaccine,* 33**,** 2015-21.

GOEBEL, W. F., BABERS, F. H. & AVERY, O. T. 1932. CHEMO-IMMUNOLOGICAL STUDIES ON CONJUGATED CARBOHYDRATE-PROTEINS : VI. THE SYNTHESIS OFp-AMINOPHENOLalpha-GLUCOSIDE AND ITS COUPLING WITH PROTEIN. *J Exp Med,* 55**,** 761-7.

GOMEZ, B., HERNANDEZ-BOU, S., GARCIA-GARCIA, J. J. & MINTEGI, S. 2014. Bacteremia in previously healthy children in Emergency Departments: clinical and microbiological characteristics and outcome. *Eur J Clin Microbiol Infect Dis*.

GRAD, Y. H., KIRKCALDY, R. D., TREES, D., DORDEL, J., HARRIS, S. R., GOLDSTEIN, E., WEINSTOCK, H., PARKHILL, J., HANAGE, W. P., BENTLEY, S. & LIPSITCH, M. 2014. Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. *The Lancet Infectious Diseases,* 14**,** 220-226.

GRAHN, E., HOLM, S. E., EKEDAHL, C. & ROOS, K. 1983. Interference of α-hemolytic streptococci isolated from tonsillar surface on β-hemolytic streptococci (Streptococcus pyogenes)—A methodological study. *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene. 1. Abt. Originale. A, Medizinische Mikrobiologie, Infektionskrankheiten und Parasitologie,* 254**,** 459-468.

GRAM, C. 1884. Ueber die isolirte Farbung der Schizomyceten in Schnitt-und Trockenpraparaten. *Fortschritte der Medicin,* 2**,** 185-189.

GRAY, B. M., CONVERSE, G. M., 3RD & DILLON, H. C., JR. 1980a. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J Infect Dis,* 142**,** 923-33.

GRAY, B. M., CONVERSE III, G. M. & DILLON JR, H. C. 1980b. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *Journal of Infectious Diseases,* 142**,** 923-933.

GREENBERG, D., GIVON-LAVI, N., BROIDES, A., BLANCOVICH, I., PELED, N. & DAGAN, R. 2006. The contribution of smoking and exposure to tobacco smoke to Streptococcus pneumoniae and Haemophilus influenzae carriage in children and their mothers. *Clinical Infectious Diseases,* 42**,** 897-903.

GRIFFITH, F. 1928. The Significance of Pneumococcal Types. *The Journal of Hygiene,* 27**,** 113-159.

GRUNSTEIN, M. & HOGNESS, D. S. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci U S A,* 72**,** 3961-5.

GSK 2017. Product information http://www.medicines.org.uk/emc/medicine/22743 accessed 24/11/17.

GWALTNEY JR, J. M., PHILLIPS, C. D., MILLER, R. D. & RIKER, D. K. 1994. Computed tomographic study of the common cold. *New England Journal of Medicine,* 330**,** 25-30.

HAKENBECK, R., GREBE, T., ZAHNER, D. & STOCK, J. B. 1999. beta-lactam resistance in Streptococcus pneumoniae: penicillin-binding proteins and non-penicillin-binding proteins. *Mol Microbiol,* 33**,** 673-8.

HALL, L. M., WHILEY, R. A., DUKE, B., GEORGE, R. C. & EFSTRATIOU, A. 1996. Genetic relatedness within and between serotypes of Streptococcus pneumoniae from the United Kingdom: analysis of multilocus enzyme electrophoresis, pulsed-field gel electrophoresis, and antimicrobial resistance patterns. *Journal of Clinical Microbiology,* 34**,** 853-9.

HANAGE, W. P., FINKELSTEIN, J. A., HUANG, S. S., PELTON, S. I., STEVENSON, A. E., KLEINMAN, K., HINRICHSEN, V. L. & FRASER, C. 2010. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. *Epidemics,* 2**,** 80-4.

HANSMAN, D. & BULLEN, M. 1967. A resistant pneumococcus. *The Lancet,* 290**,** 264-265.

HANSMAN, D., GLASGOW, H., STURT, J., DEVITT, L. & DOUGLAS, R. 1971. Increased resistance to penicillin of pneumococci isolated from man. *New England Journal of Medicine,* 284**,** 175-177.

HAQUE, F., LI, J., WU, H. C., LIANG, X. J. & GUO, P. 2013. Solid-State and Biological Nanopore for Real-Time Sensing of Single Chemical and Sequencing of DNA. *Nano Today,* 8**,** 56-74.

HARRIS, S. R., CARTWRIGHT, E. J. P., TÖRÖK, M. E., HOLDEN, M. T. G., BROWN, N. M., OGILVY-STUART, A. L., ELLINGTON, M. J., QUAIL, M. A., BENTLEY, S. D., PARKHILL, J. & PEACOCK, S. J. 2013. Whole-genome sequencing for

analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *The Lancet Infectious Diseases,* 13**,** 130-136.

HAUSDORFF, W. P., FEIKIN, D. R. & KLUGMAN, K. P. 2005. Epidemiological differences among pneumococcal serotypes. *The Lancet Infectious Diseases,* 5**,** 83-93.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.

HEFFRON, R. 1939. *Pneumonia; with special reference to pneumococcus lobar pneumonia,* New York, The Commonwealth fund; London,, H. Milford, Oxford university press.

HEIDELBERGER, M. & AVERY, O. T. 1923. The soluble specific substance of pneumococcus. *The Journal of experimental medicine,* 38**,** 73.

HILTY, M., WUTHRICH, D., SALTER, S. J., ENGEL, H., CAMPBELL, S., SA-LEAO, R., DE LENCASTRE, H., HERMANS, P., SADOWY, E., TURNER, P., CHEWAPREECHA, C., DIGGLE, M., PLUSCHKE, G., MCGEE, L., ESER, O. K., LOW, D. E., SMITH-VAUGHAN, H., ENDIMIANI, A., KUFFER, M., DUPASQUIER, M., BEAUDOING, E., WEBER, J., BRUGGMANN, R., HANAGE, W. P., PARKHILL, J., HATHAWAY, L. J., MUHLEMANN, K. & BENTLEY, S. D. 2014. Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. *Genome Biol Evol*.

HOFT, D. F., LOTTENBACH, K. R., BLAZEVIC, A., TURAN, A., BLEVINS, T. P., PACATTE, T. P., YU, Y., MITCHELL, M. C., HOFT, S. G. & BELSHE, R. B. 2017. Comparisons of the Humoral and Cellular Immune Responses Induced by Live Attenuated Influenza Vaccine and Inactivated Influenza Vaccine in Adults. *Clinical and Vaccine Immunology,* 24.

HOLLEY, R. W., APGAR, J., MERRILL, S. H. & ZUBKOFF, P. L. 1961. NUCLEOTIDE AND OLIGONUCLEOTIDE COMPOSITIONS OF THE ALANINE-, VALINE-, AND TYROSINE-ACCEPTOR "SOLUBLE" RIBONUCLEIC ACIDS OF YEAST. *Journal of the American Chemical Society,* 83**,** 4861-4862.

HORSFALL, F., GOODNER, K., MACLEOD, C. M. & HARRIS, A. 1937. Antipneumococcus rabbit serum as a therapeutic agent in lobar pneumonia. *Journal of the American Medical Association,* 108**,** 1483-1490.

HOSKINS, J., ALBORN, W. E., ARNOLD, J., BLASZCZAK, L. C., BURGETT, S., DEHOFF, B. S., ESTREM, S. T., FRITZ, L., FU, D.-J. & FULLER, W. 2001. Genome of the bacterium Streptococcus pneumoniae strain R6. *Journal of bacteriology,* 183**,** 5709-5717.

HUANG, S. S., HINRICHSEN, V. L., STEVENSON, A. E., RIFAS-SHIMAN, S. L., KLEINMAN, K., PELTON, S. I., LIPSITCH, M., HANAGE, W. P., LEE, G. M. & FINKELSTEIN, J. A. 2009. Continued Impact of Pneumococcal Conjugate Vaccine on Carriage in Young Children. *Pediatrics,* 124**,** e1-11.

HUNKAPILLER, T., KAISER, R. J., KOOP, B. F. & HOOD, L. 1991. Large-scale and automated DNA sequence determination. *Science,* 254**,** 59-67.

HUSSAIN, M., MELEGARO, A., PEBODY, R. G., GEORGE, R., EDMUNDS, W. J., TALUKDAR, R., MARTIN, S. A., EFSTRATIOU, A. & MILLER, E. 2005. A

longitudinal household study of Streptococcus pneumoniae nasopharyngeal carriage in a UK setting. *Epidemiol Infect,* 133**,** 891-8.

HYAMS, C., CAMBERLEIN, E., COHEN, J. M., BAX, K. & BROWN, J. S. 2010. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infection and immunity,* 78**,** 704-715.

HYDE, T. B., GAY, K., STEPHENS, D. S. & *ET AL*. 2001. Macrolide resistance among invasive streptococcus pneumoniae isolates. *JAMA,* 286**,** 1857-1862.

IHME 2016. History of GBD. *Institute for Health Metrics and Evaluation***,** accessed 22/11/17.

INOUYE, M., DASHNOW, H., RAVEN, L.-A., SCHULTZ, M. B., POPE, B. J., TOMITA, T., ZOBEL, J. & HOLT, K. E. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine,* 6**,** 1-16.

IVAC REPORT 2016. State of PCV Use and Impact Evaluations. *International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health*.

JACOBS, M. R., KOORNHOF, H. J., ROBINS-BROWNE, R. M., STEVENSON, C. M., VERMAAK, Z. A., FREIMAN, I., MILLER, G. B., WITCOMB, M. A., ISAÄCSON, M. & WARD, J. I. 1978. Emergence of multiply resistant pneumococci. *New England Journal of Medicine,* 299**,** 735-740.

JACOBY, P., WATSON, K., BOWMAN, J., TAYLOR, A., RILEY, T. V., SMITH, D. W., LEHMANN, D. & TEAM, K. O. M. R. P. 2007. Modelling the co-occurrence of Streptococcus pneumoniae with other bacterial and viral pathogens in the upper respiratory tract. *Vaccine,* 25**,** 2458-2464.

JAIN, M., TYSON, J. R., LOOSE, M., IP, C. L., ECCLES, D. A., O'GRADY, J., MALLA, S., LEGGETT, R. M., WALLERMAN, O. & JANSEN, H. J. 2017. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9. 0 chemistry. *F1000Research,* 6.

JANULCZYK, R., IANNELLI, F., SJÖHOLM, A. G., POZZI, G. & BJÖRCK, L. 2000. Hic, a Novel Surface Protein of Streptococcus pneumoniae That Interferes with Complement Function. *Journal of Biological Chemistry,* 275**,** 37257-37263.

JAUNEIKAITE, E., TOCHEVA, A. S., JEFFERIES, J. M. C., GLADSTONE, R. A., FAUST, S. N., CHRISTODOULIDES, M., HIBBERD, M. L. & CLARKE, S. C. 2015. Current methods for capsular typing of Streptococcus pneumoniae. *Journal of Microbiological Methods,* 113**,** 41-49.

JEDRZEJAS, M. J. 2001. Pneumococcal Virulence Factors: Structure and Function. *Microbiology and Molecular Biology Reviews,* 65**,** 187-207.

JOHNSON, S. S., ZAIKOVA, E., GOERLITZ, D. S., BAI, Y. & TIGHE, S. W. 2017. Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *Journal of Biomolecular Techniques: JBT,* 28**,** 2.

JOMBART, T., CORI, A., DIDELOT, X., CAUCHEMEZ, S., FRASER, C. & FERGUSON, N. 2014. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS computational biology,* 10**,** e1003457.

KAMERLING, J. P. 2000. Pneumococcal polysaccharides: a chemical view. *Streptococcus pneumoniae*, 81-114.

KAPATAI, G., SHEPPARD, C. L., AL-SHAHIB, A., LITT, D. J., UNDERWOOD, A. P., HARRISON, T. G. & FRY, N. K. 2016. Whole genome sequencing of Streptococcus pneumoniae: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ,* 4, e2477.

KAPATAI, G., SHEPPARD, C. L., TROXLER, L. J., LITT, D. J., FURRER, J., HILTY, M. & FRY, N. K. 2017a. Pneumococcal 23B molecular subtype identified using whole genome sequencing. *Genome Biology and Evolution*.

KAPATAI, G., SHEPPARD, C. L., TROXLER, L. J., LITT, D. J., FURRER, J., HILTY, M. & FRY, N. K. 2017b. Pneumococcal 23B Molecular Subtype Identified Using Whole Genome Sequencing. *Genome Biology and Evolution,* 9, 2122-2135.

KAUFFMAN, G. B. 1979. The discovery of penicillin: Twentieth century wonder drug. *J. Chem. Educ,* 56, 454.

KIM, J. O. & WEISER, J. N. 1998. Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of Streptococcus pneumoniae. *J Infect Dis,* 177, 368-77.

KIM, P. E., MUSHER, D. M., GLEZEN, W. P., BARRADAS, M. C. R., NAHM, W. K. & WRIGHT, C. E. 1996. Association of invasive pneumococcal disease with season, atmospheric conditions, air pollution, and the isolation of respiratory viruses. *Clinical infectious diseases,* 22, 100-106.

KING, S. J., HIPPE, K. R., GOULD, J. M., BAE, D., PETERSON, S., CLINE, R. T., FASCHING, C., JANOFF, E. N. & WEISER, J. N. 2004. Phase variable desialylation of host proteins that bind to Streptococcus pneumoniae in vivo and protect the airway. *Molecular Microbiology,* 54, 159-171.

KIRKMAN JR, J., FISCHER, J. & PAGANO, J. 1970. A microtiter plate technique for the agglutination typing of Diplococcus pneumoniae. *The Journal of infectious diseases*, 217-221.

KLAENHAMMER, T. R., BARRANGOU, R., BUCK, B. L., AZCARATE-PERIL, M. A. & ALTERMANN, E. 2005. Genomic features of lactic acid bacteria effecting bioprocessing and health. *FEMS Microbiology Reviews,* 29, 393-409.

KLUGMAN, K. 1998. Pneumococcal molecular epidemiology network. *ASM NEWS,* 64, 371-371.

KLUGMAN, K. P. 1990. Pneumococcal resistance to antibiotics. *Clinical Microbiology Reviews,* 3, 171-196.

KNUTSEN, E., JOHNSBORG, O., QUENTIN, Y., CLAVERYS, J.-P. & HÅVARSTEIN, L. S. 2006. BOX Elements Modulate Gene Expression in Streptococcus pneumoniae: Impact on the Fine-Tuning of Competence Development. *Journal of Bacteriology,* 188, 8307-8312.

KOHANSKI, M. A., DWYER, D. J. & COLLINS, J. J. 2010. How antibiotics kill bacteria: from targets to networks. *Nature reviews. Microbiology,* 8, 423-435.

KOPPE, U., SUTTORP, N. & OPITZ, B. 2012. Recognition of Streptococcus pneumoniae by the innate immune system. *Cell Microbiol,* 14**,** 460-6.

LADHANI, S. N., COLLINS, S., DJENNAD, A., SHEPPARD, C. L., BORROW, R., FRY, N. K., ANDREWS, N. J., MILLER, E. & RAMSAY, M. E. 2018. Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000-17: a prospective national observational cohort study. *The Lancet Infectious Diseases*.

LADHANI, S. N., SLACK, M. P. E., ANDREWS, N. J., WAIGHT, P. A., BORROW, R. & MILLER, E. 2013. Invasive Pneumococcal Disease after Routine Pneumococcal Conjugate Vaccination in Children, England and Wales. *Emerging Infectious Diseases,* 19**,** 61-68.

LAMMERS, A. J., DE PORTO, A. P., FLORQUIN, S., DE BOER, O. J., BOOTSMA, H. J., HERMANS, P. W. & VAN DER POLL, T. 2011. Enhanced vulnerability for Streptococcus pneumoniae sepsis during asplenia is determined by the bacterial capsule. *Immunobiology,* 216**,** 863-70.

LANIE, J. A., NG, W.-L., KAZMIERCZAK, K. M., ANDRZEJEWSKI, T. M., DAVIDSEN, T. M., WAYNE, K. J., TETTELIN, H., GLASS, J. I. & WINKLER, M. E. 2007. Genome sequence of Avery's virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6. *Journal of bacteriology,* 189**,** 38-51.

LEFEVRE, J. C., FAUCON, G., SICARD, A. M. & GASC, A. M. 1993. DNA fingerprinting of Streptococcus pneumoniae strains by pulsed-field gel electrophoresis. *Journal of Clinical Microbiology,* 31**,** 2724-2728.

LEVENE, M. J., KORLACH, J., TURNER, S. W., FOQUET, M., CRAIGHEAD, H. G. & WEBB, W. W. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science,* 299**,** 682-6.

LI, Y., CROUCHER, N. J., THOMPSON, C. M., TRZCIŃSKI, K., HANAGE, W. P. & LIPSITCH, M. 2015. Identification of pneumococcal colonization determinants in the stringent response pathway facilitated by genomic diversity. *BMC Genomics,* 16**,** 369.

LI, Y., WEINBERGER, D. M., THOMPSON, C. M., TRZCINSKI, K. & LIPSITCH, M. 2013. Surface charge of Streptococcus pneumoniae predicts serotype distribution. *Infect Immun,* 81**,** 4519-24.

LIPSITCH, M. 1997. Vaccination against colonizing bacteria with multiple serotypes. *Proceedings of the National Academy of Sciences,* 94**,** 6571-6576.

LLULL, D., GARCIA, E. & LOPEZ, R. 2001. Tts, a processive beta-glucosyltransferase of Streptococcus pneumoniae, directs the synthesis of the branched type 37 capsular polysaccharide in Pneumococcus and other gram-positive species. *J Biol Chem,* 276**,** 21053-61.

LOEFFLER, J. M. & FISCHETTI, V. A. 2006. Lysogeny of Streptococcus pneumoniae with MM1 phage: improved adherence and other phenotypic changes. *Infection and immunity,* 74**,** 4486-4495.

LOMAN, N. J., GLADSTONE, R. A., CONSTANTINIDOU, C., TOCHEVA, A. S., JEFFERIES, J. M., FAUST, S. N., O'CONNOR, L., CHAN, J., PALLEN, M. J. &

CLARKE, S. C. 2013. Clonal expansion within pneumococcal serotype 6C after use of seven-valent vaccine. *PLoS One,* 8**,** e64731.

LONG, P. H. & BLISS, E. A. 1937. Para-Amino-Benzene-Sulfonamide and Its Derivatives: Experimental and Clinical Observations on Their Use in the Treatment of Betahemolytic Streptococcic Infection: A Preliminary Report. *Journal of the American Medical Association,* 108**,** 32-37.

LOOSE, M., HUDEL, M., ZIMMER, K.-P., GARCIA, E., HAMMERSCHMIDT, S., LUCAS, R., CHAKRABORTY, T. & PILLICH, H. 2015. Pneumococcal Hydrogen Peroxide-Induced Stress Signaling Regulates Inflammatory Genes. *The Journal of Infectious Diseases,* 211**,** 306-316.

LOWELL, F. C., STRAUSS, E. & FINLAND, M. 1940. Observations on the susceptibility of pneumococci to sulfapyridine, sulfathiazole and sulfamethylthiazole. *Annals of Internal Medicine,* 14**,** 1001-1023.

MAIDEN, M. C. J., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences,* 95**,** 3140-3145.

MAJCHERCZYK, P. A. & MOREILLON, P. 2004. Inflammation and Host Defense, The Pneumococcus. ASM Press, Washington, D.C.**,** 183-200.

MANSO, A. S., CHAI, M. H., ATACK, J. M., FURI, L., CROIX, M. D. S., HAIGH, R., TRAPPETTI, C., OGUNNIYI, A. D., SHEWELL, L. K. & BOITANO, M. 2014. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nature communications,* 5**,** 5055.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARRIE, T. J. 1992. Bacteraemic pneumococcal pneumonia: a continuously evolving disease. *Journal of infection,* 24**,** 247-255.

MCGEE, L., MCDOUGAL, L., ZHOU, J., SPRATT, B. G., TENOVER, F. C., GEORGE, R., HAKENBECK, R., HRYNIEWICZ, W., LEFÉVRE, J. C., TOMASZ, A. & KLUGMAN, K. P. 2001. Nomenclature of Major Antimicrobial-Resistant Clones of Streptococcus pneumoniae Defined by the Pneumococcal Molecular Epidemiology Network. *Journal of Clinical Microbiology,* 39**,** 2565-2571.

MCKERNAN, K. J., PECKHAM, H. E., COSTA, G. L., MCLAUGHLIN, S. F., FU, Y., TSUNG, E. F., CLOUSER, C. R., DUNCAN, C., ICHIKAWA, J. K., LEE, C. C., ZHANG, Z., RANADE, S. S., DIMALANTA, E. T., HYLAND, F. C., SOKOLSKY, T.

D., ZHANG, L., SHERIDAN, A., FU, H., HENDRICKSON, C. L., LI, B., KOTLER, L., STUART, J. R., MALEK, J. A., MANNING, J. M., ANTIPOVA, A. A., PEREZ, D. S., MOORE, M. P., HAYASHIBARA, K. C., LYONS, M. R., BEAUDOIN, R. E., COLEMAN, B. E., LAPTEWICZ, M. W., SANNICANDRO, A. E., RHODES, M. D., GOTTIMUKKALA, R. K., YANG, S., BAFNA, V., BASHIR, A., MACBRIDE, A., ALKAN, C., KIDD, J. M., EICHLER, E. E., REESE, M. G., DE LA VEGA, F. M. & BLANCHARD, A. P. 2009. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res,* 19**,** 1527-41.

MELEGARO, A., CHOI, Y., PEBODY, R. & GAY, N. 2007. Pneumococcal Carriage in United Kingdom Families: Estimating Serotype-specific Transmission Parameters from Longitudinal Data.

MELEGARO, A., GAY, N. J. & MEDLEY, G. F. 2004. Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiol Infect,* 132**,** 433-41.

MILLER, E., ANDREWS, N. J., WAIGHT, P. A., SLACK, M. P. & GEORGE, R. C. 2011. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. *Lancet Infect Dis,* 11**,** 760-8.

MINA, M. J., KLUGMAN, K. P., ROSCH, J. W. & MCCULLERS, J. A. 2015. Live attenuated influenza virus increases pneumococcal translocation and persistence within the middle ear. *J Infect Dis,* 212**,** 195-201.

MINA, M. J., MCCULLERS, J. A. & KLUGMAN, K. P. 2014. Live attenuated influenza vaccine enhances colonization of Streptococcus pneumoniae and Staphylococcus aureus in mice. *MBio,* 5.

MOORE, H. F. & CHESNEY, A. M. 1917. A study of ethylhydrocuprein (optochin) in the treatment of acute lobar pneumonia. *Archives of Internal Medicine,* 19**,** 611-682.

MORGENROTH, J. & LEVY, R. 1911. *Die Chemotherapie der Pneumokokkeninfektion: 2. Mitteilung*, Hirschwald.

MOSCOSO, M., DOMENECH, M. & GARCÍA, E. 2010. Vancomycin tolerance in clinical and laboratory Streptococcus pneumoniae isolates depends on reduced enzyme activity of the major LytA autolysin or cooperation between CiaH histidine kinase and capsular polysaccharide. *Molecular Microbiology,* 77**,** 1052-1064.

MOSSER, J. & TOMASZ, A. 1970. Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *Journal of Biological Chemistry,* 245**,** 287-298.

MOSTOWY, R. J., CROUCHER, N. J., DE MAIO, N., CHEWAPREECHA, C., SALTER, S. J., TURNER, P., AANENSEN, D. M., BENTLEY, S. D., DIDELOT, X. & FRASER, C. 2017. Pneumococcal Capsule Synthesis Locus cps as Evolutionary Hotspot with Potential to Generate Novel Serotypes by Recombination. *Molecular Biology and Evolution,* 34**,** 2537-2554.

MURDOCH, J. M., SPEIRS, C., GEDDES, A. & WALLACE, E. 1964. Clinical trial of cephaloridine (Ceporin), a new broad-spectrum antibiotic derived from cephalosporin C. *British medical journal,* 2**,** 1238.

MYINT, T. T. H., MADHAVA, H., BALMER, P., CHRISTOPOULOU, D., ATTAL, S., MENEGAS, D., SPRENGER, R. & BONNET, E. 2013. The Impact of 7-valent Pneumococcal Conjugate Vaccine on Invasive Pneumococcal Disease: A Literature Review. *Advances in Therapy,* 30**,** 127-151.

NANDOSKAR, M., FERRANTE, A., BATES, E. J., HURST, N. & PATON, J. 1986. Inhibition of human monocyte respiratory burst, degranulation, phospholipid methylation and bactericidal activity by pneumolysin. *Immunology,* 59**,** 515.

NELSON, A. L., ROCHE, A. M., GOULD, J. M., CHIM, K., RATNER, A. J. & WEISER, J. N. 2007. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infection and immunity,* 75**,** 83-90.

NETTER, D. A. 1887. *De la Méningite due au pneumocoque (avec ou sans pneumonie), par le Dr Netter*, Asselin et Houzeau.

NEUFELD, F. 1902. Ueber die Agglutination der Pneumokokken und über die Theorieen der Agglutination. *Zeitschrift für Hygiene und Infektionskrankheiten,* 40**,** 54-72.

NEUFELD, F. 1910. Weiter Unterschungen Uber Pneumokokken Heilser; III. Mitleilunng Uber Vorkommen und Bedeutung Atypischer Varietaten der Pneumokokken. *Arb. adk Gsndhtsamte,* 34**,** 293-304.

NUORTI, J. P., BUTLER, J. C., FARLEY, M. M., HARRISON, L. H., MCGEER, A., KOLCZAK, M. S. & BREIMAN, R. F. 2000. Cigarette smoking and invasive pneumococcal disease. Active Bacterial Core Surveillance Team. *N Engl J Med,* 342**,** 681-9.

O'BRIEN, K. L. 2017. When less is more: how many doses of PCV are enough? *The Lancet Infectious Diseases*.

O'BRIEN, K. L., WOLFSON, L. J., WATT, J. P., HENKLE, E., DELORIA-KNOLL, M., MCCALL, N., LEE, E., MULHOLLAND, K., LEVINE, O. S., CHERIAN, T., HIB & PNEUMOCOCCAL GLOBAL BURDEN OF DISEASE STUDY, T. 2009. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet,* 374**,** 893-902.

OBERT, C., SUBLETT, J., KAUSHAL, D., HINOJOSA, E., BARTON, T., TUOMANEN, E. I. & ORIHUELA, C. J. 2006. Identification of a candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. *Infection and immunity,* 74**,** 4766-4777.

OGGIONI, M. R., MEMMI, G., MAGGI, T., CHIAVOLINI, D., IANNELLI, F. & POZZI, G. 2003. Pneumococcal zinc metalloproteinase ZmpC cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia. *Molecular microbiology,* 49**,** 795-805.

OLIVER, M. B., JONES, C., LARSON, T. R., CALIX, J. J., ZARTLER, E. R., YOTHER, J. & NAHM, M. H. 2013a. Streptococcus pneumoniae Serotype 11D Has a Bispecific Glycosyltransferase and Expresses Two Different Capsular Polysaccharide Repeating Units. *The Journal of Biological Chemistry,* 288**,** 21945-21954.

OLIVER, M. B., VAN DER LINDEN, M. P., KÜNTZEL, S. A., SAAD, J. S. & NAHM, M. H. 2013b. Discovery of Streptococcus pneumoniae serotype 6 variants with

glycosyltransferases synthesizing two differing repeating units. *Journal of Biological Chemistry,* 288**,** 25976-25985.

ORIHUELA, C. J., RADIN, J. N., SUBLETT, J. E., GAO, G., KAUSHAL, D. & TUOMANEN, E. I. 2004. Microarray Analysis of Pneumococcal Gene Expression during Invasive Disease. *Infection and Immunity,* 72**,** 5582-5596.

OWEN HENDLEY, J., SANDE, M. A., STEWART, P. M. & GWALTNEY, J. J. M. 1975. Spread of Streptococcus pneumoniae in Families. I. Carriage Rates and Distribution of Types. *The Journal of Infectious Diseases,* 132**,** 55-61.

PARK, I. H., GENO, K. A., YU, J., OLIVER, M. B., KIM, K.-H. & NAHM, M. H. 2015. Genetic, biochemical, and serological characterization of a new pneumococcal serotype, 6H, and generation of a pneumococcal strain producing three different capsular repeat units. *Clinical and Vaccine Immunology,* 22**,** 313-318.

PARK, I. H., KIM, K.-H., ANDRADE, A. L., BRILES, D. E., MCDANIEL, L. S. & NAHM, M. H. 2012. Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene pspK. *MBio,* 3**,** e00035-12.

PASTEUR, L. 1881. Note sur la maladie nouvelle provoquee par la salive d'un enfant mort de la rage. *Comptes Rendus,* 92**,** 159-165.

PÉREZ-DORADO, I., GALAN-BARTUAL, S. & HERMOSO, J. A. 2012. Pneumococcal surface proteins: when the whole is greater than the sum of its parts. *Molecular Oral Microbiology,* 27**,** 221-245.

PERICONE, C. D., OVERWEG, K., HERMANS, P. W. M. & WEISER, J. N. 2000. Inhibitory and Bactericidal Effects of Hydrogen Peroxide Production by Streptococcus pneumoniae on Other Inhabitants of the Upper Respiratory Tract. *Infection and Immunity,* 68**,** 3990-3997.

PERLINO, C. A. & RIMLAND, D. 1985. Alcoholism, Leukopenia, and Pneumococcal Sepsis 1, 2. *American Review of Respiratory Disease,* 132**,** 757-760.

PESTOVA, E. V. & MORRISON, D. A. 1998. Isolation and Characterization of Three Streptococcus pneumoniae Transformation-Specific Loci by Use of alacZ Reporter Insertion Vector. *Journal of bacteriology,* 180**,** 2701-2710.

PETROSILLO, N., PANTOSTI, A., BORDI, E., SPANO, A., DEL GROSSO, M., TALLARIDA, B. & IPPOLITO, G. 2002. Prevalence, determinants, and molecular epidemiology of Streptococcus pneumoniae isolates colonizing the nasopharynx of healthy children in Rome. *European journal of clinical microbiology & infectious diseases,* 21**,** 181-188.

PIKIS, A., CAMPOS, J. M., RODRIGUEZ, W. J. & KEITH, J. M. 2001. Optochin resistance in Streptococcus pneumoniae: mechanism, significance, and clinical implications. *The Journal of infectious diseases,* 184**,** 582-590.

PLETZ, M. W., MAUS, U., KRUG, N., WELTE, T. & LODE, H. 2008. Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. *Int J Antimicrob Agents,* 32**,** 199-206.

PMEN WEBSITE 2014. http://www.pneumogen.net/pmen/history.html. *accessed 23/11/17*.

PRICE, L. B., HUNGATE, B. A., KOCH, B. J., DAVIS, G. S. & LIU, C. M. 2017. Colonizing opportunistic pathogens (COPs): The beasts in all of us. *PLOS Pathogens,* 13**,** e1006369.

PRINCIPI, N., MARCHISIO, P., SCHITO, G. C. & MANNELLI, S. 1999. Risk factors for carriage of respiratory pathogens in the nasopharynx of healthy children. *The Pediatric infectious disease journal,* 18**,** 517-523.

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.

RADETSKY, M., JOHANSEN, T., LAUER, B., ISTRE, G., PARMELEE, S., WIESENTHAL, A. & GLODE, M. 1981. Multiply resistant pneumococcus causing meningitis: its epidemiology within a day-care centre. *The Lancet,* 318**,** 771-773.

RAMIREZ, M., SEVERINA, E. & TOMASZ, A. 1999. A high incidence of prophage carriage among natural isolates of Streptococcus pneumoniae. *Journal of bacteriology,* 181**,** 3618-3625.

RAMOS-MONTANEZ, S., KAZMIERCZAK, K. M., HENTCHEL, K. L. & WINKLER, M. E. 2010. Instability of ackA (acetate kinase) mutations and their effects on acetyl phosphate and ATP amounts in Streptococcus pneumoniae D39. *J Bacteriol,* 192**,** 6390-400.

RANDLE, E., NINIS, N. & INWALD, D. 2011. Invasive pneumococcal disease. *Archives of disease in childhood - Education & practice edition,* 96**,** 183-190.

RASKO, D. A., WEBSTER, D. R., SAHL, J. W., BASHIR, A., BOISEN, N., SCHEUTZ, F., PAXINOS, E. E., SEBRA, R., CHIN, C.-S. & ILIOPOULOS, D. 2011. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *New England Journal of Medicine,* 365**,** 709-717.

REGEV-YOCHAY, G., RAZ, M., DAGAN, R., PORAT, N., SHAINBERG, B., PINCO, E., KELLER, N. & RUBINSTEIN, E. 2004. Nasopharyngeal Carriage of Streptococcus pneumoniae by Adults and Children in Community and Family Settings. *Clinical Infectious Diseases,* 38**,** 632-639.

REGEV-YOCHAY, G., TRZCIŃSKI, K., THOMPSON, C. M., MALLEY, R. & LIPSITCH, M. 2006. Interference between Streptococcus pneumoniae and Staphylococcus aureus: in vitro hydrogen peroxide-mediated killing by Streptococcus pneumoniae. *Journal of bacteriology,* 188**,** 4996-5001.

RING, A., WEISER, J. N. & TUOMANEN, E. I. 1998. Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. *Journal of Clinical Investigation,* 102**,** 347.

RODRIGUES, F., DANON, L., MORALES-AZA, B., SIKORA, P., THORS, V., FERREIRA, M., GOULD, K., HINDS, J. & FINN, A. 2016. Pneumococcal Serotypes Colonise the Nasopharynx in Children at Different Densities. *PLOS ONE,* 11**,** e0163435.

ROMERO, P., GARCÍA, E. & MITCHELL, T. J. 2009. Development of a prophage typing system and analysis of prophage carriage in Streptococcus pneumoniae. *Applied and environmental microbiology,* 75**,** 1642-1649.

ROMERO, P., LLULL, D., GARCÍA, E., MITCHELL, T. J., LÓPEZ, R. & MOSCOSO, M. 2007. Isolation and characterization of a new plasmid pSpnP1 from a multidrug-resistant clone of Streptococcus pneumoniae. *Plasmid,* 58**,** 51-60.

ROTHBERG, J. M., HINZ, W., REARICK, T. M., SCHULTZ, J., MILESKI, W., DAVEY, M., LEAMON, J. H., JOHNSON, K., MILGREW, M. J., EDWARDS, M., HOON, J., SIMONS, J. F., MARRAN, D., MYERS, J. W., DAVIDSON, J. F., BRANTING, A., NOBILE, J. R., PUC, B. P., LIGHT, D., CLARK, T. A., HUBER, M., BRANCIFORTE, J. T., STONER, I. B., CAWLEY, S. E., LYONS, M., FU, Y., HOMER, N., SEDOVA, M., MIAO, X., REED, B., SABINA, J., FEIERSTEIN, E., SCHORN, M., ALANJARY, M., DIMALANTA, E., DRESSMAN, D., KASINSKAS, R., SOKOLSKY, T., FIDANZA, J. A., NAMSARAEV, E., MCKERNAN, K. J., WILLIAMS, A., ROTH, G. T. & BUSTILLO, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature,* 475**,** 348-52.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B. & ERLICH, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487-91.

SALUJA, S. K. & WEISER, J. N. 1995. The genetic basis of colony opacity in Streptococcus pneumoniae: evidence for the effect of box elements on the frequency of phenotypic variation. *Molecular Microbiology,* 16**,** 215-227.

SANGER, F., AIR, G. M., BARRELL, B. G., BROWN, N. L., COULSON, A. R., FIDDES, C. A., HUTCHISON, C. A., SLOCOMBE, P. M. & SMITH, M. 1977a. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature,* 265**,** 687-95.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977b. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74**,** 5463-5467.

SATZKE, C., TURNER, P., VIROLAINEN-JULKUNEN, A., ADRIAN, P. V., ANTONIO, M., HARE, K. M., HENAO-RESTREPO, A. M., LEACH, A. J., KLUGMAN, K. P., PORTER, B. D., SA-LEAO, R., SCOTT, J. A., NOHYNEK, H. & O'BRIEN, K. L. 2013. Standard method for detecting upper respiratory carriage of Streptococcus pneumoniae: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. *Vaccine,* 32**,** 165-79.

SAVE THE CHILDREN 2017. Fighting for Breath. *http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/FIGHTING-FOR-BREATH-FINAL-LOW-RES.PDF*.

SCHMIDT, L. & SESLER, C. L. 1943. Development of resistance to penicillin by pneumococci. *Proceedings of the Society for Experimental Biology and Medicine,* 52**,** 353-357.

SELANDER, R. K., CAUGANT, D. A., OCHMAN, H., MUSSER, J. M., GILMOUR, M. N. & WHITTAM, T. S. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology,* 51**,** 873-884.

SHAKHNOVICH, E. A., KING, S. J. & WEISER, J. N. 2002. Neuraminidase expressed by Streptococcus pneumoniae desialylates the lipopolysaccharide of Neisseria meningitidis and Haemophilus influenzae: a paradigm for

interbacterial competition among pathogens of the human respiratory tract. *Infection and immunity,* 70**,** 7161-7164.

SHAPER, M., HOLLINGSHEAD, S. K., BENJAMIN, W. H. & BRILES, D. E. 2004. PspA Protects Streptococcus pneumoniae from Killing by Apolactoferrin, and Antibody to PspA Enhances Killing of Pneumococci by Apolactoferrin. *Infection and Immunity,* 72**,** 5031-5040.

SHOEMAKER, N. B., SMITH, M. D. & GUILD, W. R. 1979. Organization and transfer of heterologous chloramphenicol and tetracycline resistance genes in pneumococcus. *Journal of bacteriology,* 139**,** 432-441.

SIBOLD, C., HENRICHSEN, J., KÖNIG, A., MARTIN, C., CHALKLEY, L. & HAKENBECK, R. 1994. Mosaic pbpX genes of major clones of penicillin-resistant Streptococcus pneumoniae have evolved from pbpX genes of a penicillin-sensitive Streptococcus oralis. *Molecular microbiology,* 12**,** 1013-1023.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SÖDING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* 7**,** 539-539.

SIMELL, B., AURANEN, K., KAYHTY, H., GOLDBLATT, D., DAGAN, R. & O'BRIEN, K. L. 2012. The fundamental link between pneumococcal carriage and disease. *Expert Rev Vaccines,* 11**,** 841-55.

SLATER, G. S. C. & BIRNEY, E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* [Online], 6.

SLEEMAN, K. L., GRIFFITHS, D., SHACKLEY, F., DIGGLE, L., GUPTA, S., MAIDEN, M. C., MOXON, E. R., CROOK, D. W. & PETO, T. E. A. 2006. Capsular Serotype-Specific Attack Rates and Duration of Carriage of Streptococcus pneumoniae in a Population of Children. *The Journal of Infectious Diseases,* 194**,** 682-688.

SLOYER, J. L., JR., PLOUSSARD, J. H. & HOWIE, V. M. 1981. Efficacy of pneumococcal polysaccharide vaccine in preventing acute otitis media in infants in Huntsville, Alabama. *Rev Infect Dis,* 3 Suppl**,** S119-23.

SMART, L. E. & HENRICHSEN, J. 1986. An alternative approach to typing of Streptococcus pneumoniae strains by coagglutination. *APMIS,* 94**,** 409-413.

SMITH, C. B., GOLDEN, C., KLAUBER, M. R., KANNER, R. & RENZETTI, A. 1976. Interactions between viruses and bacteria in patients with chronic bronchitis. *Journal of Infectious Diseases,* 134**,** 552-561.

SMITH, M. D. & GUILD, W. R. 1979. A plasmid in Streptococcus pneumoniae. *Journal of Bacteriology,* 137**,** 735-739.

SMITH, T., LEHMANN, D., MONTGOMERY, J., GRATTEN, M., RILEY, I. D. & ALPERS, M. P. 1993. Acquisition and invasiveness of different serotypes of Streptococcus pneumoniae in young children. *Epidemiol Infect,* 111**,** 27-39.

SPRATT, B. G. & GREENWOOD, B. M. 2000. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *The Lancet,* 356**,** 1210-1211.

STERNBERG, G. M. 1881. *A Fatal Form of Septicaemia in the Rabbit Produced by the Subcutaneous Injection of Human Saliva: An Experimental Research*, John Murphy & Company.

STERNBERG, G. M. 1885. The pneumonia-coccus of Friedlander (Micrococcus pasteuri, Sternberg). *The American Journal of the Medical Sciences,* 179**,** 106-122.

SURESH, M. V., SINGH, S. K., FERGUSON, D. A., JR. & AGRAWAL, A. 2006. Role of the property of C-reactive protein to activate the classical pathway of complement in protecting mice from pneumococcal infection. *J Immunol,* 176**,** 4369-74.

TENG, L.-J., HSUEH, P.-R., TSAI, J.-C., CHEN, P.-W., HSU, J.-C., LAI, H.-C., LEE, C.-N. & HO, S.-W. 2002. groESL sequence determination, phylogenetic analysis, and species differentiation for viridans group streptococci. *Journal of clinical microbiology,* 40**,** 3172-3178.

TETTELIN, H. & HOLLINGSHEAD, S. 2004. Comparative Genomics of *Streptococcus pneumoniae* intrastrain diversity and genome plasticity, p. 15-29. The pneumococcus. ASM Press, Washington, D.C.

TETTELIN, H., NELSON, K. E., PAULSEN, I. T., EISEN, J. A., READ, T. D., PETERSON, S., HEIDELBERG, J., DEBOY, R. T., HAFT, D. H., DODSON, R. J., DURKIN, A. S., GWINN, M., KOLONAY, J. F., NELSON, W. C., PETERSON, J. D., UMAYAM, L. A., WHITE, O., SALZBERG, S. L., LEWIS, M. R., RADUNE, D., HOLTZAPPLE, E., KHOURI, H., WOLF, A. M., UTTERBACK, T. R., HANSEN, C. L., MCDONALD, L. A., FELDBLYUM, T. V., ANGIUOLI, S., DICKINSON, T., HICKEY, E. K., HOLT, I. E., LOFTUS, B. J., YANG, F., SMITH, H. O., VENTER, J. C., DOUGHERTY, B. A., MORRISON, D. A., HOLLINGSHEAD, S. K. & FRASER, C. M. 2001a. Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. *Science,* 293**,** 498-506.

TETTELIN, H., NELSON, K. E., PAULSEN, I. T., EISEN, J. A., READ, T. D., PETERSON, S., HEIDELBERG, J., DEBOY, R. T., HAFT, D. H., DODSON, R. J., DURKIN, A. S., GWINN, M., KOLONAY, J. F., NELSON, W. C., PETERSON, J. D., UMAYAM, L. A., WHITE, O., SALZBERG, S. L., LEWIS, M. R., RADUNE, D., HOLTZAPPLE, E., KHOURI, H., WOLF, A. M., UTTERBACK, T. R., HANSEN, C. L., MCDONALD, L. A., FELDBLYUM, T. V., ANGIUOLI, S., DICKINSON, T., HICKEY, E. K., HOLT, I. E., LOFTUS, B. J., YANG, F., SMITH, H. O., VENTER, J. C., DOUGHERTY, B. A., MORRISON, D. A., HOLLINGSHEAD, S. K. & FRASER, C. M. 2001b. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science,* 293**,** 498-506.

THORS, V., CHRISTENSEN, H., MORALES-AZA, B., VIPOND, I., MUIR, P. & FINN, A. 2016a. The Effects of Live Attenuated Influenza Vaccine on Nasopharyngeal Bacteria in Healthy 2 to 4 Year Olds. A Randomized Controlled Trial. *American Journal of Respiratory and Critical Care Medicine,* 193**,** 1401-1409.

THORS, V., CHRISTENSEN, H., MORALES-AZA, B., VIPOND, I., MUIR, P. & FINN, A. 2016b. The Effects of Live Attenuated Influenza Vaccine on

Nasopharyngeal Bacteria in Healthy 2 to 4 Year Olds. A Randomized Controlled Trial. *Am J Respir Crit Care Med,* 193**,** 1401-9.

TILLETT, W. S., CAMBIER, M. J. & MCCORMACK, J. E. 1944. The Treatment of Lobar Pneumonia and Pneumococcal Empyema with Penicillin. *Bull N Y Acad Med,* 20**,** 142-78.

TILLEY, S. J., ORLOVA, E. V., GILBERT, R. J., ANDREW, P. W. & SAIBIL, H. R. 2005. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell,* 121**,** 247-256.

TOCHEVA, A. S., JEFFERIES, J. M., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2010. Increase in serotype 6C pneumococcal carriage, United Kingdom. *Emerg Infect Dis,* 16**,** 154-5.

TOCHEVA, A. S., JEFFERIES, J. M., RUBERY, H., BENNETT, J., AFIMEKE, G., GARLAND, J., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2011a. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of Streptococcus pneumoniae in young children. *Vaccine,* 29**,** 4400-4.

TOCHEVA, A. S., JEFFERIES, J. M. C., RUBERY, H., BENNETT, J., AFIMEKE, G., GARLAND, J., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2011b. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of Streptococcus pneumoniae in young children. *Vaccine,* 29**,** 4400-4404.

TOMASZ, A. & MOSSER, J. 1966. On the nature of the pneumococcal activator substance. *Proceedings of the National Academy of Sciences,* 55**,** 58-66.

TREANGEN, T. J., ONDOV, B. D., KOREN, S. & PHILLIPPY, A. M. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome biology,* 15**,** 524.

TUOMANEN, E., LIU, H., HENGSTLER, B., ZAK, O. & TOMASZ, A. 1985. The induction of meningeal inflammation by components of the pneumococcal cell wall. *Journal of Infectious Diseases,* 151**,** 859-868.

TURNER, P., HINDS, J., TURNER, C., JANKHOT, A., GOULD, K., BENTLEY, S. D., NOSTEN, F. & GOLDBLATT, D. 2011. Improved Detection of Nasopharyngeal Cocolonization by Multiple Pneumococcal Serotypes by Use of Latex Agglutination or Molecular Serotyping by Microarray. *Journal of Clinical Microbiology,* 49**,** 1784-1789.

VAMMEN, B. 1939. Serological Variants of Pneumoeoeeus Types 9 and 10. *Journal of Immunology,* 37**,** 359-65.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. & THERMES, C. 2014. Ten years of next-generation sequencing technology. *Trends Genet,* 30**,** 418-26.

VAN HOEK, A. J., SHEPPARD, C. L., ANDREWS, N. J., WAIGHT, P. A., SLACK, M. P. E., HARRISON, T. G., LADHANI, S. N. & MILLER, E. 2014. Pneumococcal carriage in children and adults two years after introduction of the thirteen valent pneumococcal conjugate vaccine in England. *Vaccine,* 32**,** 4349-4355.

VAN OPIJNEN, T., BODI, K. L. & CAMILLI, A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods,* 6**,** 767-772.

VAN SELM, S., VAN CANN, L. M., KOLKMAN, M. A., VAN DER ZEIJST, B. A. & VAN PUTTEN, J. P. 2003a. Genetic basis for the structural difference between Streptococcus pneumoniae serotype 15B and 15C capsular polysaccharides. *Infect Immun,* 71**,** 6192-8.

VAN SELM, S., VAN CANN, L. M., KOLKMAN, M. A. B., VAN DER ZEIJST, B. A. M. & VAN PUTTEN, J. P. M. 2003b. Genetic Basis for the Structural Difference between Streptococcus pneumoniae Serotype 15B and 15C Capsular Polysaccharides. *Infection and Immunity,* 71**,** 6192-6198.

VAN TONDER, A. J., BRAY, J. E., JOLLEY, K. A., QUIRK, S. J., HARALDSSON, G., MAIDEN, M. C. J., BENTLEY, S. D., HARALDSSON, A., ERLENDSDOTTIR, H., KRISTINSSON, K. G. & BRUEGGEMANN, A. B. 2017. Heterogeneity Among Estimates Of The Core Genome And Pan-Genome In Different Pneumococcal Populations. *bioRxiv*.

VESIKARI, T., WYSOCKI, J., CHEVALLIER, B., KARVONEN, A., CZAJKA, H., ARSÈNE, J.-P., LOMMEL, P., DIEUSSAERT, I. & SCHUERMAN, L. 2009. Immunogenicity of the 10-valent pneumococcal non-typeable Haemophilus influenzae protein D conjugate vaccine (PHiD-CV) compared to the licensed 7vCRM vaccine. *The Pediatric infectious disease journal,* 28**,** S66-S76.

WAIGHT, P. A., ANDREWS, N. J., LADHANI, S. N., SHEPPARD, C. L., SLACK, M. P. & MILLER, E. 2015. Effect of the 13-valent pneumococcal conjugate vaccine on invasive pneumococcal disease in England and Wales 4 years after its introduction: an observational cohort study. *Lancet Infect Dis,* 15**,** 535-43.

WALTER, N. D., TAYLOR, T. H., DOWELL, S. F., MATHIS, S. & MOORE, M. R. 2009. Holiday Spikes in Pneumococcal Disease among Older Adults. *New England Journal of Medicine,* 361**,** 2584-2585.

WALTER , N. D., TAYLOR , T. H. J., DOWELL , S. F., MATHIS , S. & MOORE , M. R. 2009. Holiday Spikes in Pneumococcal Disease among Older Adults. *New England Journal of Medicine,* 361**,** 2584-2585.

WANNAMAKER, L. W. & MATSEN, J. M. 1972. *Streptococci and streptococcal diseases: Recognition, Understanding, and Management.*

WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids. *Nature,* 171**,** 737-738.

WEICHSELBAUM, A. 1886. Aetiologie und pathologische Anatomie der akuten lungenentzundungen. *Wiener Medizinishe Wochenschrift,* 36**,** 1301-5.

WEINBERGER, D. M., GRANT, L. R., WEATHERHOLTZ, R. C., WARREN, J. L., O'BRIEN, K. L. & HAMMITT, L. L. 2016. Relating Pneumococcal Carriage Among Children to Disease Rates Among Adults Before and After the Introduction of Conjugate Vaccines. *American Journal of Epidemiology,* 183**,** 1055-1062.

WEISER, J. 2010. The pneumococcus: why a commensal misbehaves. *Journal of Molecular Medicine,* 88**,** 97-102.

WEISER, J. N., BAE, D., FASCHING, C., SCAMURRA, R. W., RATNER, A. J. & JANOFF, E. N. 2003. Antibody-enhanced pneumococcal adherence requires IgA1 protease. *Proceedings of the National Academy of Sciences,* 100**,** 4215-4220.

WHITE, B. 1938. The Biology of the Pneumococcus. *The Commonwealth Fund, New York, N.Y.***,** p. XVI-XVII.

WHITNEY, C. G., GOLDBLATT, D. & O'BRIEN, K. L. 2014. Dosing schedules for pneumococcal conjugate vaccine: considerations for policy makers. *The Pediatric infectious disease journal,* 33**,** S172.

WYRES, K. L., VAN TONDER, A., LAMBERTSEN, L. M., HAKENBECK, R., PARKHILL, J., BENTLEY, S. D. & BRUEGGEMANN, A. B. 2013. Evidence of antimicrobial resistance-conferring genetic elements among pneumococci isolated prior to 1974. *BMC Genomics,* 14**,** 500-500.

YESILKAYA, H., ANDISI, V. F., ANDREW, P. W. & BIJLSMA, J. J. E. 2013. Streptococcus pneumoniae and reactive oxygen species: an unusual approach to living with radicals. *Trends in Microbiology,* 21**,** 187-195.

YESILKAYA, H., SPISSU, F., CARVALHO, S. M., TERRA, V. S., HOMER, K. A., BENISTY, R., PORAT, N., NEVES, A. R. & ANDREW, P. W. 2009. Pyruvate Formate Lyase Is Required for Pneumococcal Fermentative Metabolism and Virulence. *Infection and Immunity,* 77**,** 5418-5427.

YOTHER, J. 2011. Capsules of Streptococcus pneumoniae and other bacteria: paradigms for polysaccharide biosynthesis and regulation. *Annu Rev Microbiol,* 65**,** 563-81.

YPMA, R. J. F., VAN BALLEGOOIJEN, W. M. & WALLINGA, J. 2013. Relating Phylogenetic Trees to Transmission Trees of Infectious Disease Outbreaks. *Genetics,* 195**,** 1055-1062.

ZAUFAL, E. 1887. Mikroorganismen im Secrete der Otitis media acuta. *Prager Medicinische Wochenschrift,* 12**,** 225-7.