Pneumococcal carriage and disease amongst children from the United Kingdom and Nepal

Dr Rama Kandasamy

St. Cross College
University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of *Doctor of Philosophy*

Trinity 2017

Supervisors:
Professor Andrew J. Pollard, Department of Paediatrics, University of Oxford, Oxford, Oxfordshire.

Doctor Dominic F. Kelly, Department of Paediatrics, University of Oxford, Oxford, Oxfordshire.

Professor Stephen Bentley, Department of Pathogen Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire.

Abstract

Pneumococcal disease continues to be responsible for an extraordinary amount of human death and disability around the world. In particular, the burden of disease weighs greatest upon those children who reside in low and low-middle income countries of the world. It must also be recognised however, that the continued emergence of disease due to pneumococcal serotypes not covered by pneumococcal conjugate vaccines, due to serotype replacement, is an unfolding issue which is increasingly impacting upon the health of children in middle and high income countries.

In the short and medium term work needs to be focused on achieving the greatest benefits from the intelligent use of currently available pneumococcal conjugate vaccines. Whilst the long term vision towards combating pneumococcal disease needs to focus on better understanding the underlying biology, particularly through the use of contemporary technologies.

As an insight into the above mentioned issues faced by resource-rich countries, this thesis aims to describe the dynamics of pneumococcal carriage and disease in the United Kingdom following the introduction of the thirteen-valent pneumococcal conjugate vaccine (PCV13) to the infant immunisation schedule.

From the perspective of a resource limited setting, this thesis also aims to utilise the population of pneumococci collected from amongst Nepalese children prior to PCV introduction to; determine the pneumococcal serotype distribution, determine the utility of a molecular diagnostic tool for identifying pneumococcal pneumonia, determine the molecular epidemiology of pneumococci in this population, and determine the antibiotic resistance patterns of pneumococci in this population.

Finally, with the intention of exploring the overall diversity of pneumococcus and how this may be influenced by PCV introduction, this thesis aims to also describe the global population structure of pneumococci prior to PCV introduction.

Key findings of this thesis include; the differential effect of PCV13 on serotypes 3 and 19A, and the rise of non-vaccine type disease amongst children from the United Kingdom, the characterisation of carriage prior to PCV10 introduction and the high rate of multiple serotype carriage amongst Nepalese children, the identification of host immune and antibiotic selective pressures on pneumococcal surface protein A (pspA), pneumococcal surface protein C (pspC), and penicillin binding proteins, the almost unchecked rise of antimicrobial resistance amongst pneumococci in Nepal, and the description of the global pneumococcal population prior to PCV introduction, which shows that some strains have characteristics which facilitate spread across geographic regions.

These findings highlight the need for further studies investigating how PCVs could be better applied in United Kingdom infants and children in order to try and minimise the observed differential effects. The identification of the pneumococcal surface proteins which are under selective pressure (pspA and pspC) highlights these proteins as promising antigens for inclusion in a novel vaccine that would provide an avenue for preventing pneumococcal disease using a non-serotype specific approach. The resistance patterns of pneumococci in Nepal indicate that further steps both from a surveillance and policy stand point need to be taken to monitor and curb pneumococcal antimicrobial resistance in Nepal. Finally, it is clear that ongoing surveillance of pneumococcal carriage and disease is needed both in the United Kingdom and Nepal. Of most interest will be the post-PCV effect on the pneumococcal population in Nepal.

Acknowledgements

The work described in this thesis is the result of a large body of work and collaboration between the Oxford Vaccine Group, the Patan Academy of Health Sciences, and the Wellcome Trust Sanger Institute (WTSI).

I am extremely grateful to Andrew Pollard for bringing together all of the amazing opportunities for me to explore throughout these last three years. He has given very kind support throughout the various phases of the projects I have worked upon. Most importantly he has been always approachable and available, both as a supervisor and as a person who genuinely cares about others, even when we are on opposite sides of the world. I would also like to thank Dominic Kelly for his tireless support and seemingly endless depth of knowledge and thought on any given topic.

There are many staff at the Oxford Vaccine Group who have been involved in the conduct of the studies that are included in this thesis and to who I am indebted. I would like to thank Sarah Kelly for all of her programmatic management which has kept all of the studies in Nepal running throughout some of the most challenging times. I would like to warmly thank Michael Carter for being a beacon of calmness in the hectic atmosphere of Kathmandu. Thanks also must go to Catrin Moore and Susana Camara for all of their help with processing samples and help with setting up my assays.

The Patan Academy of Health Sciences has had over a decade long collaboration with the Oxford Vaccine Group and as such there have been numerous people who have contributed to the recruitment of participants and collection of samples for the projects that are described in this thesis. Importantly thanks must go to Stephen Thorson and Meeru Gurung for all of the tireless work they have put into running the projects in Nepal. I must also thank Madhav Gautam for all of his work serotyping pneumococci.

The Wellcome Trust Sanger Institute leads the world in the investigation of all things at a molecular level. And it was here I was fortunate enough to spend some time learning about pathogen genomics. Here I must thank Ste Bentley, Rebecca Gladstone, John Lees, and Nicholas Croucher for all of the very kind support and guidance they provided in opening up the world of bioinformatics to someone such as myself (a clinician). If it wasn't for the rich network of the WTSI I would never have met Jukka Corander when he was on sabbatical from Helsinki and who I must thank for running the global population clustering on his own personal server, a task that would not have been computationally achievable otherwise.

Finally, I would like to give acknowledgement those to whom this thesis existence would not have been realised without them. To Christoph Blöhmke for his unwavering and tireless support, at any hour, on any day. For being so understanding and seemingly always able to condense everything down into the most logical thought. To Kristina Lunz for giving me the light and inspiration to pursue my work over these last three years, beyond what would ever have been possible otherwise. For it is without such love I would never have had the will to do all that has followed on the pages herein. And to my Mum and Dad for everything that has lead me up to this point.

Declaration and contributions

The research presented in this thesis is original work with any contributions not made by the author clearly stated in the text. This thesis includes data generated from studies lead by the Oxford Vaccine Group, the Patan Academy of Health Sciences, and the Wellcome Trust Sanger Institute. The author acknowledges the work of the clinical study staff at the Oxford Vaccine Group and the Patan Academy of Health Sciences for the collection and processing of samples. Whole genome sequencing of pneumococcal isolates was conducted by the core sequencing team at the Wellcome Trust Sanger Institute and was funded by the Bill and Melinda Gates Foundation. Within the time-period over which the projects for this thesis were conducted the author was awarded the 2016 Robert Austrian Research Award in Pneumococcal Vaccinology at the 10th International Symposium on Pneumococci and Pneumococcal Diseases (ISPPD). This award is financially supported by Pfizer limited, however it is independently awarded by the ISPPD awards committee.

Publications

Published

R. Kandasamy, M. Voysey, F. McQuaid, K. de Nie, R. Ryan, O. Orr, U. Uhlig, C. Sande, D. O'Connor, and A.J. Pollard. Non-specific immunological effects of selected routine childhood immunisations: systematic review. *BMJ*. 2016 Oct 13;355:i5225.

M. Voysey, **R. Kandasamy**, L.M. Yu, M. Baudin, C. Sadorge, S. Thomas, T. John, and A.J. Pollard. The predicted persistence and kinetics of antibody decline 9 years after pre-school booster vaccination in UK children. *Vaccine*. 2016 Jul 29;34(35):4221-8.

M. Hamaluba*, **R. Kandasamy***, N. Adhikari, S. Thorson, S. Upreti, S. Shrestha, S. Bhattarai, Gurung, R. Pradhan, M. Voysey, S. Gurung, A. Habib, S. Pradhan, A. K. Thapa, R. Maharjan, U. Kiran, J. Habens, S. Dewan, S. A. Kerridge, J. Metz, J. Hinds, A. Werno, F. van der Klis, M. D. Snape, D. R. Murdoch, S. Kelly, D. F. Kelly, and A. J. Pollard. A randomised controlled open-label immunogenicity study of the 10-Valent Pneumococcal Conjugate Vaccine; comparing two dose priming with booster to a three dose priming schedule in Nepalese infants. *Lancet Infectious Disease* 2015 Apr;15(4):405-14. ***joint first authors**

R. Kandasamy, M. Gurung, A. Thapa, S. Ndimah, N. Adhikari, D. R. Murdoch, D. F. Kelly, D. E. Waldron, K. A. Gould, S. Thorson, S. Shrestha, J. Hinds, A. J. Pollard. Multi-serotype pneumococcal nasopharyngeal carriage prevalence in vaccine naïve Nepalese children, assessed using molecular serotyping. *PLoS One*. 2015 Feb 2;10(2)

M. Hamaluba, **R. Kandasamy**, S. Ndimah, R. Morton, M. Caccamo, H. Robinson, S. Kelly, A. Field, L. Norman, E. Plested, B. A. Thompson, A. Zafar, S. A. Kerridge, R. Lazarus, T. John, J. Holmes, S. N. Fenlon, K. A. Gould, P. Waight, J. Hinds, D. Crook, M. D. Snape, and A. J. Pollard. A cross-sectional observational study of pneumococcal carriage in children, their parents, and older adults following the introduction of the 7-valent pneumococcal conjugate vaccine. *Medicine* (Baltimore). 2015 Jan;94(1)

R. Kandasamy, F. McQuaid, M. Voysey, K. de Njie, O. Orr, R. Ryan, D. O'Connor, C. Sande, and A. J. Pollard. Systematic Review of the Non-specific Immunological Effects of Selected Routine Childhood Immunizations. World Health Organization, Department of Immunization, Vaccines, and Biologicals, SAGE. April 2014.

Submitted or in preparation

R. Kandasamy, M. Voysey, S. Collins, G. Berbers, H. Robinson, I. Noel, H. Hughes, S. Ndimah, N. Fry, S. Ladhani, M. D. Snape, J. Hinds, A. J. Pollard. Persistent circulation of vaccine serotypes and serotype replacement with high attack-rate non-vaccine types after five years of UK infant immunisation with PCV13. *Submitted to JAMA Paediatrics (Peer reviewed by PLoS Medicine)*

E. Balsells, R. Dagan, I. Yildirim, P. P. Gounder, A. Steens, C. Muñoz-Almagro, C. Mameli, **R. Kandasamy**, N. G. Lavi, L. Daprai, A. van der Ende, K. Trzciński, S. Nzenze, S. Meiring, D. Foster, L. Bulkow, K. Rudolph, A. Valero-Rello, S. Ducker, D. Frimann Vestrheim, A. von Gottberg, S. Pelton, G. Zuccotti, A. J. Pollard, E. Sanders, H. Campbell, S. A. Madhi, H. Nair,

M. H. Kyaw. *Streptococcus pneumoniae* serotype-specific invasive disease potential after PCV introduction: a systematic review and meta-analysis. *Submitted to Lancet Infectious Disease*

M. Voysey, T. R Fanshawe, D. F Kelly, K. L. O'Brien, **R. Kandasamy**, S. Shrestha, S. Thorson, J. Hinds, A. J. Pollard. Serotype-specific correlates of protection for pneumococcal carriage: an analysis of immunity in 5224 infants in 19 countries. *Submitted to Lancet Infectious Disease*

R. Kandasamy, D. F. Kelly, O. Ramilo, M. Natea, C. L. Karp, C. B. Wilson, P. Fine, A. J. Pollard. Heterologous immunological effects of childhood vaccination in the 21st century.

Table of contents

List of figures

List of tables

Glossary

AUC	Area under the curve
BAPS	Bayesian analysis of population structure
BC	BAPS cluster
BLAST	Basic local alignment search tool
CDC	Centre for Disease Control
CDS	Coding DNA sequence
CI	Confidence interval
CSF	Cerebrospinal fluid
DNA	Deoxyribose nucleic acid
EPC	End-point consolidation
EUCAST	European Committee on Antimicrobial Susceptibility Testing
GPS	Global Pneumococcal Sequencing (project)
IPD	Invasive pneumococcal disease
IQR	Inter-quartile range
MLST	Multi-locus sequence typing
NPNM	Non-pneumonia, non-meningitis pneumococcal disease
NT	Non-typeable
NVT	Non-vaccine type
PCR	Polymerase chain reaction
PCV	Pneumococcal conjugate vaccine
PCV7	Seven-valent pneumococcal conjugate vaccine
PCV10	Ten-valent pneumococcal conjugate vaccine
PCV13	Thirteen-valent pneumococcal conjugate vaccine
QC	Quality control

qPCR	Quantitative polymerase chain reaction
rRNA	Ribosomal ribonucleic acid
SD	Standard deviation
SDPs	Silica desiccant packages
SNP	Single nucleotide polymorphism
STGG	Skim-milk-tryptone-glucose-glycerine
UR	Uncertainty range
UTI	Urinary tract infection
VRT	Vaccine related type
WHO	World Health Organization
WTSI	Wellcome Trust Sanger Institute

Chapter 1 Introduction

Streptococcus pneumoniae (pneumococcus) is a Gram-positive coccus that is often carried asymptomatically in the nasopharynx, but may cause illnesses ranging in severity from skin infections and otitis media to pneumonia, septicaemia, and meningitis. Primarily through the causation of pneumonia, pneumococcal disease is a major cause of child mortality globally, with peak incidence in Africa and Asia. Following wide-spread use in resource-rich settings pneumococcal conjugate vaccines (PCVs) are being introduced into resource-limited settings, however there are on-going challenges to disease prevention which include: variable vaccine impact, achieving optimal vaccine efficacy, understanding how the bacterium evolves in the context of PCVs, and accurately diagnosing the disease.

1.1 Microbiology of pneumococcus

1.1.1 The discovery of pneumococcus

Pneumococcus was first isolated by George Miller Sternberg in 1880, following the injection of his own saliva into a rabbit which subsequently died of sepsis.[1] Louis Pasteur similarly isolated the bacterium following the injection of saliva from an infant, who died of rabies, into a rabbit, which appeared to sicken and die much more quickly than would be expected by the typically slow progression of rabies.[2] Pasteur and Sternberg both published their findings in 1881. During the ensuing decade the application of Hans Christian Gram's staining method, by himself and others, to samples from patients who had died from lobar pneumonia led to the attribution of pneumococcus to being the most common cause of community acquired pneumonia at the time.[3] It was then subsequently discovered that certain complications associated with pneumonia, such as endocarditis, septic arthritis, and meningitis were the result of the haematogenous spread of pneumococci from the primary infection in the respiratory tract to distant sites.[4–6]

In 1886 G. and F. Klemperers described transferable immunity using serum from animals that had recovered from infection.[7] This paved the way for Neufeld to describe the serological agglutination and swelling (Quellung reaction) of pneumococcus using serum from infected animals, which later in collaboration with Haendel led to the realisation that more than one strain of pneumococcus existed.[8–10]

1.1.2 Culture conditions

Pneumococcus may occur as single cocci, dipplococci, or in lancet shaped forms. It is a fastidious bacterium and grows best at 35-37°C in five percent carbon dioxide. Typical culture conditions for pneumococcus are a nutrient agar supplemented with blood which provides a source of catalase that hydrolyses peroxides, protecting the bacteria from oxidative damage.[11] Colonies grown on blood agar are typically small, grey, and moist in appearance with an associated zone of dark green alpha-haemolysis. Pneumococcus may be differentiated from closely related *Streptococcus spp.* of similar appearance by testing optochin (a quinine analogue) sensitivity and bile solubility.[12]

1.1.3 The capsule

Pneumococcus typically produces a polysaccharide capsule covering the cell wall, which serves as a method of evading host defences by inhibiting complement activity and preventing phagocytic uptake.[13] Some strains do not express a capsule, and are referred to as non-typeable pneumococci. The first serologically distinct capsular types (serotypes 1 and 2) of pneumococcus were described in 1910 by Neufeld and Haendel with at least 92 structurally and serologically distinct variations (serotypes) now described.[10,14] Two different systems of serotype nomenclature have been developed, the Danish and the American systems, with the Danish system now being the most mainstream system used in scientific reporting. The Danish

system groups together cross-reactive types into common serogroups and then sub-divides these into the respective serotypes, whereas the American system numbers each serotype by order of discovery.[15]

The thickness of the capsule ranges from 200-400 nm, and varies according to serotype.[16] Generally, the capsule polysaccharide is covalently bound to the peptidoglycan of the cell wall, although this is reportedly not the case for serotype 3.[17] The polysaccharide structure varies from simple linear chains to more complex branching chains of monosaccharides.[18]

1.1.4 Cell wall composition

The cell wall is comprised chiefly of teichoic acid and peptidoglycan. The peptidoglycan is comprised of *N*-acetylglucosamine, *N*-acetylmuramic acid, and a lysine containing stem peptide.[19] There are three main groups of pneumococcal cell surface proteins which anchor to the cell wall. Namely, choline binding proteins, LPTXG anchored proteins (which are anchored to the cell wall with a LPTXG amino acid motif), and lipoproteins.[20]

1.1.5 Virulence factors

In addition to the polysaccharide capsule, pneumococci possess a battery of virulence factors, with some of the key factors highlighted in table 1. Other than bacteriocins, which are targeted at other bacteria, pneumococcal virulence factors work to either counteract components of the host immune system or to bind to host cells.

Table 1. ***Key pneumococcal virulence factors and described functions***

Virulence factor	Function/s
Pneumococcal surface protein A (pspA)	Binds lactoferrin, thus reducing the antibacterial activity of lactoferrin.[21] Prevents binding of complement component C3, thus inhibiting opsonisation.[22]
Pneumococcal surface protein C (pspC)	Binds to complement component C3, complement factor H, poly immunoglobulin receptor, and IgA secretory fragment.[23–26]
Autolysin (lytA)	Breaks down peptidoglycan in the pneumococcal cell wall resulting in release of pneumolysin.[27]
Pneumolysin (ply)	Forms transmembrane pores in host cells resulting in lysis.[28]
Neuraminidase	Neuraminidases function to cleave sialic acid from glycoproteins, glycolipids and oligosaccharides in the mucous and on the host cell surface.[29]

1.1.6 Phase variation

Two morphological phases of pneumococci, namely transparent and opaque, have been described, with strains of pneumococci often being observed to be growing as a mixture of these phenotypes.[30] One initial study demonstrated the transparent phase as being advantageous for nasopharyngeal colonisation in a rat model.[31] However in the same study there were no differences between the two phases with respect to bacteraemia or lethality when given by intraperitoneal injection. The later finding may be attributed to the possibility that the strains used were able to change phase following injection, as the finding is in contrast to a more recent study in mice. A recent study identified a pneumococcal genomic locus which when modified also altered the phase. The authors then developed strains with fixed versions of this locus, and hence the bacteria were unable to alter phase. Subsequent challenge of mice with these strains demonstrated that the transparent phase strains were better at maintaining persistent colonisation whilst the opaque phase strains were better at persisting in the blood.[32]

1.2 Pathogenesis

Pneumococcal disease can manifest in many forms, but is most commonly seen in the context of upper respiratory tract infections, pneumonia, bacteraemia, and meningitis. Pneumococcal disease typically begins after the bacteria spread from the nasopharynx, where it initially

resides asymptomatically, to other bodily sites. Localised infection of the upper respiratory tract occurs when pneumococcus enters into the sinuses or the Eustachian tube causing sinusitis and otitis media respectively. Pneumonia results when bacteria in the nasopharynx are aspirated into the lower respiratory tract, with studies suggesting that viral co-infection may facilitate this process.[33,34] The presence of pneumococci in the alveoli stimulates an acute inflammatory response, with the alternate complement pathway facilitating neutrophilic phagocytosis of the bacteria.[35] The surrounding polysaccharide capsule of pneumococcus inhibits phagocytosis, whilst other virulence factors also target and attempt to disable host immune defence mechanisms, including steps in the complement pathways.[36] Inflammation within the lung tissue may lead to a fluid effusion in the pleural space, with reports of between 40-57% of people hospitalised with suspected pneumococcal pneumonia having an effusion detected.[37] Infection in the pleural space, known as empyema may develop, although it is less common and only reportedly occurs in 5-10% of those with a pleural effusion.[38] Pneumococci have surface proteins which can bind to host surface proteins present on epithelial and endothelial cells, binding of these host receptors activates transportation pathways which act to internalise the bacteria attached at the apical surface and transport it to the basal surface, a process known as transcytosis.[23,39] From the bloodstream pneumococci may disseminate to distant sites, most commonly of which is the meninges.[40] In the case of meningeal invasion, studies have shown that the same receptor mediated translocation processes involved in migrating from the respiratory epithelium to the blood are involved in crossing the blood-brain barrier.[26]

1.3 Pneumococcal disease diagnostics

1.3.1 Nasopharyngeal carriage

Nasopharyngeal colonisation is the prelude to pneumococcal disease. However nasopharyngeal colonisation itself is a poor diagnostic tool for pneumonia and invasive disease, due to the fact that many children who have lower respiratory tract infections or invasive disease due to other pathogens may also be colonised, and a large proportion of healthy children carry pneumococcus asymptomatically.

1.3.2 Classical microbiology

The traditional approach of direct microbiological isolation of pneumococcus from a sterile site such as blood or pleural aspirate lacks sensitivity for diagnosing pneumococcal pneumonia. For blood cultures this is largely because pneumococcal pneumonia does not always result in bacteraemia and thus detection of pneumococcus in blood only serves a diagnostic utility when bacteraemic pneumonia is present. This is exemplified in adults where a meta-analysis of studies reporting blood culture results in patients with community acquired pneumonia estimated that bacteraemia could be detected by blood culture in only 25% of pneumococcal pneumonia cases.[41] In children the yield is even lower with studies reporting less than 10% of children with pneumococcal pneumonia having a detectable organism by blood culture.[42,43]

Culture of pleural fluid aspirate provides a method for identifying cases of non-bacteraemic pneumococcal pneumonia. Studies have shown however, that conventional microbiological assessment of pleural lung aspirates varies considerably in the ability to detect an infectious aetiology.[44–46] Additionally, the procedure has intrinsic risks associated with it, such as pneumothorax, is not standard clinical practice for the majority of pneumonia cases, and rarely affects the course of clinical treatment.

1.3.3 PCR

Molecular techniques such as PCR detection of pneumococcal DNA have been utilised in a bid to increase the yield of sterile site fluid analysis, particularly in the circumstance where antibiotics have been pre-administered. A systematic review examining PCR for the detection of pneumococcal DNA in blood of children with any assay (culture, serology, or antigen testing) proven IPD (including pneumonia) compared with healthy children and/or non-IPD proven disease had a sensitivity of 41.7% and specificity of 98.7%.[47]

Studies examining PCR analysis of pleural fluids from children with empyema, where pleural fluid sampling is usually standard practice, show a significantly higher ability to detect pneumococcus (70-75%) compared with conventional culture (11-40%).[48,49]

1.3.4 Urine antigen detection

Studies of adults with pneumonia which have used a technique for detecting pneumococcal polysaccharides in urine using a multiplex immunoassay as a diagnostic test have recently been reported.[50,51] The assay detects only the thirteen polysaccharides which are contained within PCV13 and thus is targeted towards assessing vaccine efficacy rather than diagnosing pneumococcal pneumonia. Given this limitation the assay does however have an impressive sensitivity of 97% and specificity of 100% for detecting the presence of pneumococcus in adult patients with bacteraemic pneumococcal pneumonia due to one of the PCV13 serotypes.[50]

An alternative technique which detects urinary pneumococcal antigen by immuno-chromatography, has been developed and commercialised.[52] The antigen of detection is the pneumococcal C-polysaccharide (PnC) that is present on the cell surface of all pneumococci. This technique has particular advantages around the ease of sample collection and its utility post antibiotic therapy. Meta-analysis of adult studies has shown that this test has a specificity of 79.6% and sensitivity 76.7% in those compared with a gold standard of blood culture

positive pneumococcal pneumonia.[53] Meta-analysis which included studies using a composite diagnostic test, combining sputum and other sterile site cultures, reported a specificity of 84.2% and a modestly reduced sensitivity of 68.5%.[53]

However, these measures are less impressive in children, who have an inherently higher asymptomatic carriage rate compared to adults. Specifically in children, asymptomatic nasopharyngeal carriage appears to confound the assays detection ability as one study shows that 42.5% of healthy children who carry pneumococcus had a positive urinary antigen test, whilst another study reported 54% of children without respiratory tract infections compared with 61% of children with pneumonia, who all carried pneumococcus, tested positive for pneumococcal urinary antigen.[54,55] This study showed that there was a strong association for a positive urine antigen test and nasopharyngeal colonisation irrespective of the disease process.

1.3.5 Serology

Serological testing for immunoglobulins to pneumococcal antigens has yielded very little with regard to a clinically useful assay to identify pneumococcal pneumonia in children. The assays have been primarily based on detecting antibody responses to pneumolysin, PnC, and serotype specific capsular motifs. A paper reviewing the literature on this topic demonstrated these assays to be positive in 27-38% of children with radiologically confirmed pneumonia.[56] A number of very small studies examining the correlation between blood culture and serology were also reviewed in this paper and demonstrated a wide range (47-92%) of agreement between the tests. In adults, larger studies have demonstrated a significant lack in the utility of these assays for diagnosing pneumococcal disease, due to the high levels of baseline and healthy control immunogenicity to the assay antigens.[57,58] These assays continue to serve a use only as a research tool due to poor sensitivity and specificity, technical challenges, and the time required to confirm seroconversion.

1.3.6 A role for new approaches and new technologies

The inherent difficulty in diagnosing pneumococcal pneumonia using mainstream approaches has lead researchers to apply new methodologies to the problem. The application of next generation molecular techniques such as transcriptomics has been posited, however this approach in general is yet to yield any clear diagnostic tool for diseases which are comparatively easier to define. Studies in adults have looked at the use of measuring the density of pneumococcal nasopharyngeal carriage via PCR and found a sensitivity of 82.2% and a specificity of 92.0% for distinguishing community acquired pneumococcal pneumonia (as measured using a composite diagnostic tool) from asymptomatic colonisation.[59] This assay has been used in children with pneumonia where it was found that carriage density of pneumococcus correlates with severe pneumonia, however it has not been used in children with pneumonia as a diagnostic aid, and appears to be a promising avenue for exploration.[60]

1.4 The evolution of pneumococcal conjugate vaccines

The first experiments assessing a pneumococcal vaccine were in 1911 and led by Almroth Wright whose expertise was enlisted in an attempt to curb the high rate of pneumonia observed in South African miners.[61] These experiments used heat killed pneumococci as the inoculum which was given to a large number of miners. It was inconclusive however, whether this vaccine had any effect on disease, with a decrease in pneumonia observed, but no reduction in overall disease related deaths. Subsequently Frederick Lister, through the appreciation of the serological diversity of pneumococcus and the refinement of Wrights' experiments, which had used too few bacteria in the inoculum, conducted trials of multi-valent whole-cell vaccines in miners. Unfortunately there was a lack of efficacy of the vaccines, which was probably due to the mismatching of strains contained in the vaccine with those that were causing disease in the mining camps.[62]

It was in the 1930's that Francis and Tillet demonstrated the immunogenicity of the capsular polysaccharide following the injection of serotype-specific polysaccharides into adults.[63] This finding led to field trials, which coupled with the advent of the second world war resulted in large trials being conducted amongst U.S.A. troops. These trials led to the licensing of two hexa-valent polysaccharide vaccines in the 1940's in the U.S.A.[64] This was soon followed by the discovery of penicillin which quenched further vaccine development in the coming decades.

In the 1970's the observation that the peak of pneumococcal disease in the community was amongst infants and young children facilitated the need to develop a vaccine that was effective in these populations. Higher valency pneumococcal polysaccharide vaccines were also being developed, however the immunogenicity of pneumococcal polysaccharides in children was poor. This barrier was only later overcome following the development of a highly immunogenic vaccine in children using *Haemophilus influenzae* polysaccharide conjugated to

bacterial exotoxins such as tetanus or diphtheria toxoids which then led to the application of exotoxin conjugation to pneumococcal polysaccharides.[65] A seven-valent vaccine which was later followed by a 13-valent vaccine, where the polysaccharides are conjugated to the altered diphtheria toxin CRM is produced by Pfizer limited. Similarly, a 10-valent vaccine with all of the polysaccharides except 18C (conjugated to tetanus toxoid) and 19F (conjugated to diphtheria toxoid) conjugated to *Haemophilus influenzae* protein D is manufactured by GlaxoSmithKline (table 2). New PCVs are now currently in development and include; a 10-valent PCV produced by the Serum Institute of India, a 15-valent PCV produced by Merck, and a 20-valent PCV produced by Pfizer.[66,67]

Table 2. ***Constituents of pneumococcal conjugate vaccines.***

Vaccine (trade name, manufacturer)	Conjugate protein (amount, µg)	Serotypes covered (amount of polysaccharide, µg)												
PCV7 (Prevnar 7, Pfizer)	CRM_{197} (20)	4 (2)	6B (4)	9V (2)	14 (2)	18C (2)	19F (2)	23F (2)						
PCV10 (Synflorix, GlaxoSmithKline)	*H. influenzae* protein D for serotypes 1, 4, 5, 6B, 7F, 9V, 14, and 23F (9-16). Tetanus toxoid for serotype 18C (5-10). Diphtheria toxoid for serotype 19F (3-6).	4 (3)	6B (1)	9V (1)	14 (1)	18C (3)	19F (3)	23F (1)	**1 (1)**	**5 (1)**	**7F (1)**			
PCV13 (Prevnar 13, Pfizer)	CRM_{197} (34)	4 (2.2)	6B (4.4)	9V (2.2)	14 (2.2)	18C (2.2)	19F (2.2)	23F (2.2)	**1 (2.2)**	**5 (2.2)**	**7F (2.2)**	**3 (2.2)**	**6A (2.2)**	**19A (2.2)**

All vaccines contain aluminium phosphate as an adjuvant, PCV7=0.125 mg, PCV10=0.5 mg, and PCV13=0.125 mg.[68]

CRM_{197}= Non-toxic variant of diphtheria toxin isolated from cultures of *Corynebacterium diphtheriae* strain C7 (β197).

1.5 Global and regional burden of pneumococcal disease in childhood, with a focus on Europe and South Asia

1.5.1 Worldwide, pneumonia and meningitis are responsible for a large proportion of childhood deaths

It has been estimated that in 2013, global all-cause child mortality was 3,665,700.[69] Of these, lower respiratory infections formed the largest proportion of communicable causes, 1,768,800, whilst meningitis is estimated to be responsible for 262,500 deaths. Importantly these estimations also demonstrate disease burden is not equally distributed across all regions, with resource-limited regions, such as South Asia being most heavily affected. Of these all-cause estimates, pneumococcus specifically plays a leading role in the causation of pneumonia, meningitis, and other forms of invasive bacterial disease in children worldwide.

1.5.2 The burden of pneumococcal pneumonia

The lack of a reliable diagnostic tool for determining pneumococcal pneumonia makes it intrinsically difficult to make precise estimations of its disease burden. However, through the modelling of data from large vaccine efficacy trials (vaccine probe approach), conducted in the U.S.A., South Africa, the Gambia, and the Philippines, pneumococcal pneumonia was estimated to have a global incidence of 2,228 (95% UR 1,733-2,772) per 100,000 and be responsible for 741,000 (95% UR 542,000-805,000) deaths in children aged under 5 years in 2000.[70–74]

At a regional level Southeast Asia is estimated to have a pneumococcal pneumonia incidence of 2911 (95% UR 2265-3622) per 100,000 with a mortality of 169,000 (95% UR 124,000–184,000) in children under 5 years. Contrastingly in Europe, prior to PCV introduction, pneumococcal pneumonia had an estimated incidence of 462 (95% UR 359-574) per 100,000 with a corresponding mortality of 13,000 (95% UR 9,500-14,200).

1.5.3 The burden of pneumococcal meningitis

As children with suspected meningitis routinely have cerebrospinal fluid collected for microbiological analysis, more reliable estimates of pneumococcal meningitic disease burden can be made based on surveillance data. Data from a WHO surveillance report for 2013, indicates that compared with *Haemophilus influenzae* and *Neisseria meningititis*, pneumococcus continues to be the most frequent cause (81.7%) of bacterial meningitis amongst Southeast Asia region children less than 5 years of age.[75] Estimations for the year 2000 indicated an incidence of 13 (UR 11-16) per 100,000 and that there were 24,200 (95% UR 19,500-28,500) cases of pneumococcal meningitis in the Southeast Asia region of whom there was a case fatality rate of 57% (95% UR 44-71%).[70]

Although pneumococcal conjugate vaccine introduction is yet to be comprehensively established in these settings it would be expected that more recent estimations would show a reduction in incidence as access to health care and improvements in living circumstances have improved.

In Europe the pre-PCV estimate of incidence for pneumococcal meningitis for the year 2000 among children under 5 years was 6 (UR 5-9) per 100,000 with a case fatality rate of 38% (UR 32-58%).[76] There is currently no update on the incidence of pneumococcal meningitis in this age group across Europe following PCV introduction. However, country-specific estimates indicate large decreases in pneumococcal meningitis. For example, in the UK a decrease was seen from 3.18 cases per 100,000 prior to PCV7 introduction to 1.44 cases per 100,000 following PCV7 introduction in children under 5 years of age.[77]

1.5.4 The burden of non-pneumonia, non-meningitis pneumococcal disease

It is well recognised that pneumococcus may establish disease in bodily sites other than the lungs and cause invasive disease other than meningitis (for example bacteraemia), that is; non-pneumonia, non-meningitis pneumococcal disease (NPNM). In resource-limited areas it is intrinsically difficult to estimate the NPNM disease burden as most cases probably occur outside the hospital setting. However, estimates have been generated based on a small set of data describing proportions of meningitis and pneumonia to NPNM disease. From these calculations, in Southeast Asia it is estimated that NPNM disease had an incidence of 67 per 100,000 totaling 122,000 cases in the year 2000. In Europe the equivalent incidence was 36 per 100,000 resulting in 15,200 cases over the year.[70]

1.5.5 Pneumococcal disease in the United Kingdom and Nepal: Specific examples of pneumococcal disease in resource-rich and resource-limited settings.

1.5.5.1 Resource-rich and resource-limited countries each face their own challenges to pneumococcal disease control and provide unique opportunities for understanding disease mechanisms

Although the resource-limited countries bear the bulk of the pneumococcal disease burden it is the resource-rich countries which have the most mature infant PCV programmes. Countries with mature PCV programmes need to contend with issues of optimising vaccine efficacy and new strain emergence. Whilst those countries that are in the process of introducing a PCV to their programme face challenges posed by logistics, cost-effectiveness, and the alternate patterns of disease when compared with other settings. Obviously resource-limited countries will benefit from the evidence generated by studies and observations made following PCV implementation in resource-rich settings, for example reduced dose schedules. Though it is important to also realise that those settings which are only beginning to implement a PCV

provide a unique opportunity to study vaccine effect using contemporary technologies which weren't accessible during the initial introduction of PCVs into resource-rich countries.

In resource-rich countries the most important intervention in the last decade for combating pneumococcal disease has been the introduction of PCVs. PCV7 was licensed in Europe in 2001, with most countries implementing the vaccine into the infant immunisation schedule between 2006 and 2008.[78,79] In resource-limited countries the last decade has seen dramatic improvements in general living standards and access to healthcare. Most notably the almost unrestricted access to antimicrobial medications in some settings is likely to have had a role in at least lessening the severity of some pneumococcal disease, but not without the consequential rise of antimicrobial resistance. The now ensuing introduction of PCVs into African and Asian countries with high pneumococcal disease burden marks the start of a pivotal period for disease prevention in these settings.

1.5.5.2 Comparison of the UK to Nepal at a population and fiscal level

In 2015 the UK had an estimated population of 65.1 million people of whom about 81% reside in urban areas.[80,81] The UK is considered a wealthy country, having the worlds 5th largest economy, and ranks 37th in the world for GDP per capita. It is estimated that over £116 billion was spent on the provision of public healthcare in the UK in the 2015/2016 fiscal year.[82]

Nepal is a landlocked country in South Asia, being bordered by China to the north and India to the south, east and west. Nepal had an estimated population of 26.5 million in 2011 of whom 17% resided in an urban setting.[83] Based on GDP per capita in 2015, Nepal is one of the poorest countries in the world, ranking 200th of 229 countries.[84] Health expenditure in the 2013/2014 fiscal year was estimated to be $US0.3 billion.[85]

1.5.5.3 Pneumococcal pneumonia in UK children

Changes in the number of children hospitalised with lower respiratory tract infections (pneumonia and empyema) in the UK provides an insight into the effect of PCVs on these illnesses. It is important however, to consider the following points regarding these data. Firstly, other than the introduction of PCVs, there have not been any major changes or introduction of vaccines against bacterial causes of pneumonia since *Haemophilus influenzae* vaccination in 1992. So it would not be expected that there are any large changes to all-cause pneumonia incidence due to vaccines against non-pneumococcal bacterial pathogens over the last two decades. Secondly, variations in incidence of viral causes of pneumonia, such as differences in influenza disease burden, exemplified by the H1N1 influenza A outbreak in 2009, may have confounding effects on extrapolating all-cause pneumonia data to bacterial causes. Finally, there are some confounders which are specific to how the data is collected, for example a retrospective review of a hospital admissions database may be confounded by practices around coding diagnoses.

In the five years leading up to the introduction of PCV7, observational hospital admission data from Scotland and England show a progressive increase in all-cause pneumonia across all ages.[86,87] Interestingly, there is then a decrease from 2006 onwards, which coincides with PCV7 introduction. In both data sets, the majority of paediatric pneumonia admissions are in children under 2 years, and it is in this age group that there is the greatest decline in admissions from 2005 to 2006. These changes seem to occur too soon to be attributed to the effect of PCV7, and it is possible that they are due either to viral induced disease or fluctuation in causative bacterial strains. For example, serotype 1 pneumococcal disease typically occurs as outbreaks.[88] Such a hypothesis for a bacterial cause is supported by the fact that the English data also show a decrease in incidence of empyema.

Following the trough in incidence in 2006 there is then a gradual increase in incidence of pneumonia admissions in both countries which peaks in 2010 around the time of introduction of PCV13. The Scottish data then shows a decline in paediatric pneumonia admissions, however the English data show no change compared with the pre-PCV7 era. Interestingly the English data do show a decline in admissions for empyema in children under 2 years when comparing the pre-PCV13 with the post-PCV13 era. Thus it was hypothesised by the authors that this observation may be because the additional serotypes covered by PCV13 are more likely to be associated with complicated pneumonia and/or IPD.[89]

1.5.5.4 Serotype-specific distribution of childhood invasive pneumococcal disease in the United Kingdom

IPD surveillance across England and Wales is now centralised, with Public Health England commencing enhanced surveillance of IPD cases in 2010, coinciding with the introduction of PCV13 the same year. Prior to 2010, Public Health England had conducted surveillance of IPD isolates from England and Wales which were voluntarily submitted to its reference laboratory, whilst some UK regional centres reported IPD data independently. So it is important to note that the pre-PCV13 data reported by Public Health England does not completely represent the total incidence across England and Wales. As such the publications from Public Health England reporting IPD data have adjusted the raw values for changes in reporting. Using the available published data, the serotype-specific distribution of IPD in children, with a focus on PCV13 covered serotypes is subsequently detailed according to the three time periods, pre-PCV7, post-PCV7 (pre-PCV13), and post-PCV13.

1.5.5.4.1 Serotype-specific IPD in UK children prior to PCV7 introduction.

Before the introduction of PCV7 there are three paediatric cohorts from the UK that have had serotype-specific IPD data reported (table 3). There are a further five cohorts reporting IPD data from UK children however they are from earlier time periods when data were reported by serogroup rather than serotype. Across the three more recently reported cohorts the estimated coverage of IPD by PCV7 was consistent, ranging from 75.4-77% (table 4). The most frequent IPD causing serotype covered by a PCV was serotype 14.

1.5.5.4.2 Serotype-specific IPD in UK children after PCV7 introduction but before PCV13 introduction.

Post PCV7 nationwide serotype-specific IPD case data for England and Wales are available for two time periods.[77,90,91] One of these time-periods is reported twice with slightly differing values, due to additional statistical adjustments made in the more recent report.[90] Miller *et al* reports significant reductions in overall PCV7 type IPD in all paediatric age groups when compared with the incidence of PCV7 type IPD in the pre-PCV7 period spanning July 2000 to June 2006. The estimated coverage of PCV13, ranging from 66.1-67.7%, is consistent between the two reported cohorts.

Miller *et al* compares IPD incidence across England and Wales from 2000-2006 (pre-PCV7) with 2008-2010 (post-PCV7) at a serotype-specific level and demonstrates a significant decline for each of the PCV7 covered serotypes in children under 5 years when compared with the incidence of each serotype in the pre-PCV7 period. A significant decrease was also seen for serotype 6A/C for this age group. Interestingly this study also reported findings consistent with serotype replacement with significant increases in IPD incidence for children under 5 years for non-vaccine types 3, 7F, 19A, 10A, 15C, 22F, and 23B.

1.5.5.4.3 Serotype-specific IPD in UK children after PCV13 introduction

Post PCV13 nationwide serotype-specific IPD case data for England and Wales are available for the time period from July 2013 to June 2014 and were compared to those from the July 2008 to June 2010 period described above.[90] Comparing the two time periods, the data show significant declines in overall IPD incidence due to the additional six serotypes covered by PCV13 for all of the paediatric age groups. A continued decline in the overall incidence of IPD due to PCV7 covered serotypes is also shown. When comparing the incidence of IPD in children under 5 years from the pre-PCV13 to the post-PCV13 era at a serotype-specific level there was a significant decline for the PCV13 serotypes 1, 6A, 7F, and 19A. Interestingly one paper reported a significant increase in incidence for the non-vaccine type 24F.

Table 3. ***Characteristics of UK studies reporting serotype-specific data on invasive pneumococcal disease isolates from children for PCV13 serotypes.***

First author (publication year)	Sleeman (2001)	Foster (2008)	Miller (2011)	Miller (2011)	Waight (2015)	Hamaluba (2015)	Waight (2015)
Time period of collection	July 1995-December 1999	January 1996-December 2005	July 2000-June 2006	July 2008-June 2010	July 2008-June 2010	November 2010-September 2011	July 2013-June 2014
Age group	<5 years	<5 years	<5 years	<5 years	<5 years	<5 years	<5 years
Region	Oxfordshire	Oxfordshire	England and Wales	England and Wales	England and Wales	England and Wales	England and Wales
Vaccine period	Pre-PCV7	Pre-PCV7	Pre-PCV7	Post-PCV7	Post-PCV7	Post-PCV13	Post-PCV13
Total number of isolates	138	446	1071	528	461	98	247

White columns = Pre-PCV7, green columns = Post-PCV7, and blue columns = Post-PCV13.

Table 4. ***Serotype-specific data on invasive pneumococcal disease isolates from UK children for PCV13 serotypes.***

Serotype	Sleeman (2001), n (cases/year)	Foster (2008)[‡], n (cases/year)	Miller (2011)[#], n (cases/year)	Miller (2011)[#], n (cases/year)	Waight (2015)[f], n (cases/year)	Hamaluba (2015), n (cases/year)	Waight (2015)[f], n (cases/year)
4	4 (0.7)	10 (1)	36 (6)	3 (1.5)	NR	-	-
6B	11 (2)	54 (5.4)	134 (22.3)	8 (4)	NR	-	-
9V	11 (2)	23 (2.3)	57 (9.5)	2 (1)	NR	-	-
14	37 (6.7)	125 (12.5)	310 (51.7)	3 (1.5)	NR	-	-
18C	22 (4)	50 (5)	84 (14)	7 (3.5)	NR	-	-
19F	12 (2.2)	45 (4.5)	115 (19.2)	15 (7.5)	NR	1 (1.2)	-
23F	7 (1.3)	38 (3.8)	73 (12.2)	4 (2)	NR	-	-
1	6 (1.1)	12 (1.2)	47 (7.8)	68 (34)	59 (29.5)	23 (27.6)	5 (5)
5	-	-	3 (0.5)	5 (2.5)	-	-	-
7F	5 (0.9)	-	24 (4)	104 (52)	90 (45)	20 (24)	8 (8)
3	4 (0.7)	-	17 (2.8)	30 (15)	26 (13)	5 (6)	8 (8)
6A	7 (1.3)	17 (1.7)	40* (6.7)	19 (9.5)	10 (5)	-	-
19A	7 (1.3)	18 (1.8)	45 (7.5)	100 (50)	85 (42.5)	16 (19.2)	7 (7)
% PCV7	75.4	77	75.5	8	8.2	1	3.2
% PCV13	96.4	87.9	88.2	66.1	67.7	66.3	15

[‡]This study only reported the 10 most frequent serotypes for the age group. [#]Number of cases reported were adjusted for missing data, changes in population, and changes in trends of acquisition. *This value is for both serotypes 6A and 6C which were not distinguished by the diagnostic test used. [f]Number of cases reported were adjusted for the number of samples serotyped, missing age, and trend for IPD up to 2010. Serotype 5 was deemed an outbreak strain and corrected cases for each epidemiological year from 2000/2001 to 2013/2014 were 49, 5, 7, 23, 5, 0, 16, 62, 43, 10, 4, 1, 4, 0. NR=Not reported. White columns = Pre-PCV7, green columns = Post-PCV7, and blue columns = Post-PCV13.

1.5.5.5 Serotype-specific distribution of invasive pneumococcal disease in Nepalese children

There is no nation-wide pneumococcal disease surveillance programme in Nepal, however there are data published from two hospital based (Kanti Children's Hospital and Patan Hospital) surveillance programmes in Kathmandu, Nepal (table 5). Both of these sites demonstrate a large disease burden of paediatric pneumonia with Patan Hospital reporting 753 cases in a 21-month period (for children aged up to 12 years) and Kanti Children's Hospital reporting 2069 cases over a 31-month period (for children aged 2 months - 5 years). In the same time periods, for the same age ranges, there were 700 and 243 cases of suspected meningitis in Patan and Kanti Children's hospitals respectively. At Patan Hospital *Salmonella* Typhi was the most commonly isolated organism overall, however pneumococcus was the most commonly isolated organism in children under one year of age. At Kanti Children's Hospital pneumococcus was the most frequently isolated organism.

When examining the serotype-specific IPD cases in Kathmandu, PCV7 covered serotypes were isolated infrequently (5.9-14.3% of all isolates), whilst the additional serotypes covered by PCV10, represent a much higher proportion of disease (35.3-73.2%) (table 6). The majority of disease is attributed to serotypes 1 and 5, which are covered by PCV10. There was infrequent disease due to the additional three serotypes covered by PCV13 that are not covered by PCV10.

The data from the Patan Hospital surveillance programme, presented by *Gurung et al*, indicates two important patterns of paediatric IPD at the Patan Hospital site (figure 1). Firstly, there appears to be a waxing and waning of IPD incidence, which is attributed to outbreaks of serotype 1 disease. Secondly, disease frequency peaks in early infancy and the majority of this disease is due to serotypes which are not covered by PCV10. With the peak of PCV10 covered disease occurring in late infancy and extending into early childhood.

Table 5. ***Characteristics of studies reporting serotype-specific data on invasive pneumococcal disease isolates from Nepalese children.***

First author (publication year)	Shah (2009)	Rijal (2010)	Williams (2009)	Kelly (2011)	Gurung (2016)
Study period	November 2004 - March 2007	November 2004 - December 2008	April 2005 - December 2006	April 2005 - December 2006	April 2005 - November 2015
Age group	2 months - 5 years	2 months - 5 years	2 months - 5 years	≤12 years	0-14 years
Region	Kanti, Children's Hospital, Kathmandu	Kanti, Children's Hospital, Kathmandu	Patan Hospital, Kathmandu	Patan Hospital, Kathmandu	Patan Hospital, Kathmandu
Number of cases	49	60	17	22	123

Table 6. ***Nepal studies reporting serotype-specific data on invasive pneumococcal disease isolates from children for the 13 PCV serotypes.***

Serotype	Shah (2009), n (cases/year)	Rijal (2010), n (cases/year)	Williams (2009), n (cases/year)	Kelly (2011), n (cases/year)	Gurung (2016), n (cases/year)
4	-	-	-	-	2 (0.2)
6B	1 (0.4)	1 (0.2)	-	-	2 (0.2)
9V	-	-	-	2 (1.1)	2 (0.2)
14	1 (0.4)	2 (0.5)	-	-	5 (0.4)
18C	-	-	1 (0.6)	1 (0.6)	1 (0.1)
19F	1 (0.4)	1 (0.2)	-	-	2 (0.2)
23F	1 (0.4)	2 (0.5)	-	-	2 (0.2)
1	8 (3.3)	13 (3.2)	4 (2.3)	10 (5.7)	50 (4.4)
5	3 (1.2)	9 (2.2)	1 (0.6)	2 (1.1)	14 (1.2)
7F	2 (0.8)	2 (0.5)	-	-	1 (0.1)
3	-	-	-	-	1 (0.1)
6A	-	-	-	-	1 (0.1)
19A	-	-	1 (0.6)	1 (0.6)	2 (0.2)
PCV7 %	14.3	13	5.9	8.8	14.3
PCV10 %	60.7	65.2	35.3	44.1	73.2
PCV13 %	60.7	65.2	41.2	47.1	75.9

Serotypes coloured black are those covered by PCV7. Serotypes coloured green are those additional serotypes covered by PCV10 and PCV13. Serotypes coloured red are those additional serotypes covered by PCV13 only.

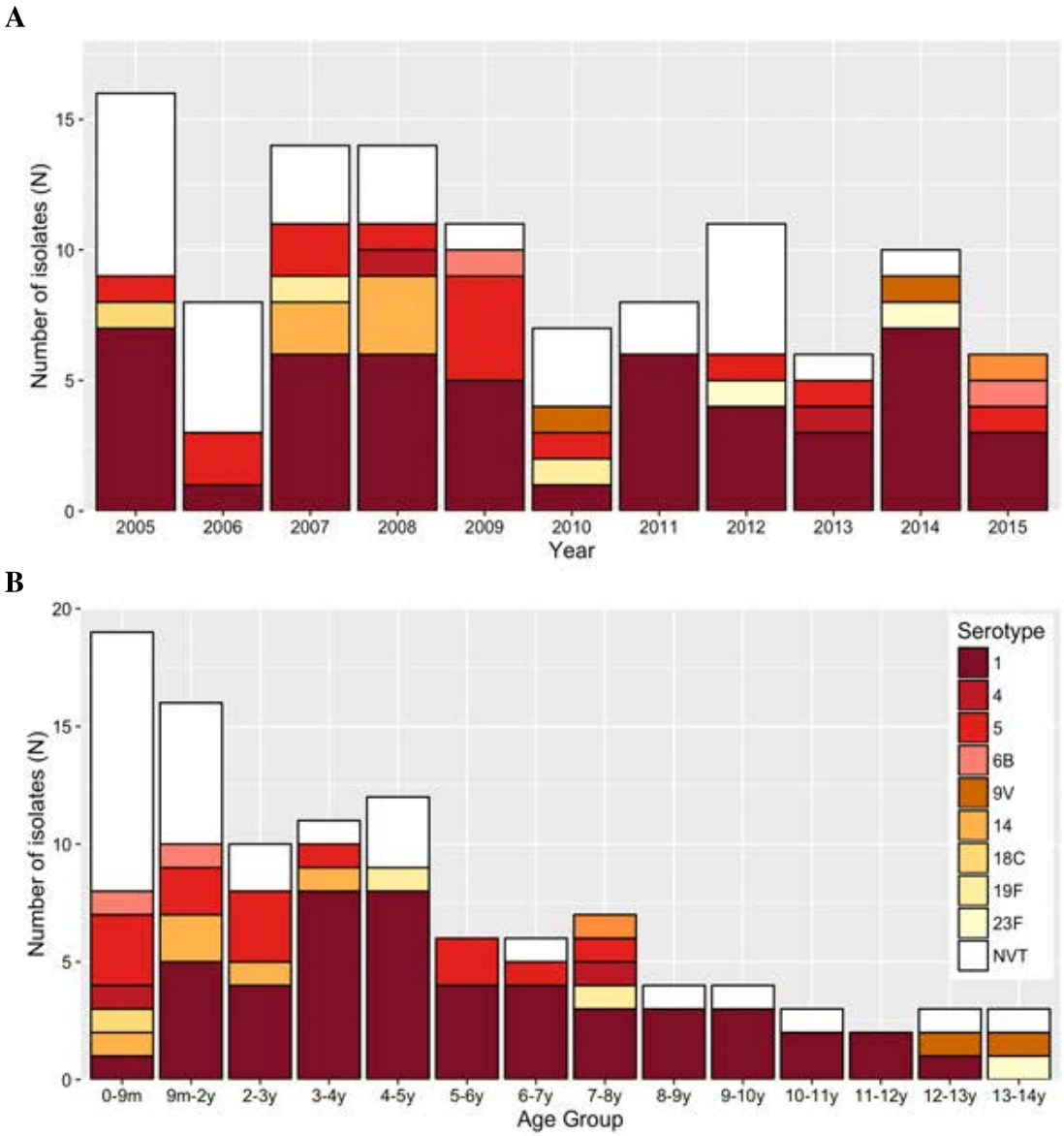


Figure 1. ***Serotype-specific frequency of IPD in Nepalese children aged 0-14 years attending Patan Hospital, Kathmandu between April 2005 and November 2015 according to year (A) and age-group (B).***
A total of 123 patients had pneumococcus detected in a sterile culture, of which 112 were serotyped. Bars are shaded according to the frequency of each of the PCV10 serotypes or as non-vaccine types (NVT). No serotype 7F isolates were detected.

1.6 The use of nasopharyngeal carriage to assess PCV effect

1.6.1 The rationale for conducting nasopharyngeal carriage studies

The use of PCVs in different regions where there are different patterns of disease has brought about a need to assess vaccine effect across these different populations. Exemplified by the fact that 74 countries to date have ongoing or published PCV impact studies.[92] However, although pneumococcus is frequently carried in the nasopharynx of humans asymptomatically and transmitted via airborne droplets. Only a very small proportion of people develop invasive disease following exposure, with a combination of bacterial, host, and environmental factors at play. As such the distribution of invasive disease may not necessarily reflect the transmission pattern of pneumococcus. Furthermore, large, well-funded, government supported, nation-wide surveillance programmes are needed in order to acquire enough invasive disease case numbers to inform vaccine use and impact in real-time. In comparison with large scale invasive disease surveillance, nasopharyngeal carriage is assessed by non-invasive means and can be inexpensively performed across large cohorts, providing a more feasible means for evaluating PCV effect, epidemiology, and transmission patterns. Carriage studies are also able to identify risk factors for transmission and when combined with IPD data can provide insight into the invasiveness (the odds that a given serotype has of being carried compared with causing disease) of specific serotypes. The application of new technologies to the assessment of carriage, for example DNA microarray assays, which can not only detect multiple pneumococcal serotypes that may be present within a single sample, but also antibiotic resistance genes, and closely related non-pneumococcal *Streptococcus spp.*, provide an exciting opportunity to explore the biology of carriage at a molecular level.

1.6.2 Risk factors for carriage

A variety of factors have been implicated in altering the risk of pneumococcal carriage. Few however are modifiable and thus provide limited opportunity for interrupting transmission. Specifically, these factors include, age, sex, exposure to smoke, attendance at facilities with close quarters contact or living, recent antibiotic consumption, and seasonality.

1.6.2.1 The relationship of age to pneumococcal carriage prevalence

The number of children asymptomatically carrying pneumococcus progressively increases throughout infancy and into early childhood, with peak carriage prevalence in the toddler and pre-school years.[93] Carriage levels progressively decline into adolescence and adulthood, and are lowest in older adults.[91,94]

1.6.2.2 The relationship of sex to pneumococcal carriage prevalence

No studies have reported female sex as a risk factor for carriage however, two studies have shown that male sex is a risk factor for carriage.[94,95] It should be noted however, that both of the studies showing a higher risk of carriage in males were conducted in adolescent and adult age groups, and it has been hypothesised that other factors such as alcohol consumption and smoke exposure may be the true underlying factors influencing these observations.

1.6.2.3 Increased carriage prevalence is seen in children attending daycare or pre-schools

Respiratory droplet transmission coupled with a peak carriage prevalence in toddler and pre-school aged children is the probable reason as to why attendance at institutional settings such as day-care centres and pre-schools where children can closely interact has been associated with increased carriage prevalence.[96] In addition, it also explains why paediatric carriage

studies have also shown that having more children in the household or having an older sibling can be associated with a higher risk of carriage.[97,98]

1.6.2.4 Children exposed to open fires have an increased risk of pneumococcal carriage

Conceptually smoke may impair the mucociliary clearance of bacteria from the airways, whilst it might also irritate the airways and facilitate transmission of bacteria through coughing and sneezing. In resource-limited regions exposure to smoke from open fires has been shown to be a risk factor for carriage in children,[95,99] however it is less clear as to whether exposure to passive cigarette smoke in resource-rich settings confers any increased risk of carriage, with no studies in children under five years of age identifying such a relationship.[91,98]

1.6.2.5 Antibiotic consumption decreases the risk of pneumococcal carriage

Recent antibiotic consumption of participants needs to be considered when pneumococcal carriage is being assessed because it has been shown to reduce carriage prevalence. For example, a longitudinal study of community based children has shown that carriage prevalence is lower in children following a course of antibiotics.[100] Whilst sampling of children with respiratory tract infections before and after receipt of antibiotics showed a significant decrease in carriage prevalence following antibiotic treatment.[101]

1.6.2.6 Carriage prevalence is higher during winter months than in summer months

The prevalence of pneumococcal carriage in children has been shown to have seasonal variations, with higher prevalence during the winter months and lower prevalence in the summer months.[102] This observation is likely to be related to increased viral illnesses during the winter months which would facilitate airborne transmission through coughing and sneezing.

1.6.3 The frequency of carriage compared with disease can be used to identify invasive strains

Longitudinal carriage studies show that different serotypes tend to colonise children for different lengths of time, for example, in one study serotype 6B was reported to have an average carriage duration of 19.9 weeks, whilst serotype 7F had a carriage duration of 5.5 weeks.[97] The strains that have long durations of carriage are those that are most frequently identified in cross-sectional carriage studies, whilst strains with short durations are infrequently found.

When carriage prevalence is related to invasive disease cases at a serotype specific level it becomes apparent that some serotypes are infrequently found in carriage but are frequently found in the context of invasive disease, that is the strains are highly invasive.[97,103]

Many carriage studies have shown that overall carriage rates remain relatively unchanged following vaccine introduction.[91,104,105] This is because NVTs occupy the population space left by the vaccine types. Some of these NVTs may be highly invasive and become more prevalent in disease, as was seen following the introduction of PCV7 in the UK. Monitoring for the emergence of highly invasive strains should continue to be a priority in the post-PCV setting. This is because development of new pneumococcal conjugate vaccines would have the greatest impact on disease if they were targeted towards highly invasive strains rather than strains of lower invasiveness.

1.6.4 Serotype-specific distribution of pneumococcal carriage amongst children in the United Kingdom: an example of PCV effect measured through carriage.

Pneumococcal carriage studies have been conducted on children in the United Kingdom to assess the effect of both PCV7 and PCV13. Six papers, reporting twelve study cohorts, were found to have extractable data on serotype-specific pneumococcal carriage in UK children (table 7). Of these, two study cohorts performed sample collections prior to PCV7 introduction,

five following PCV7 introduction, and five after PCV13 introduction. It is important to note that the two cohorts studied in the pre-PCV7 period were longitudinal studies, where participants were swabbed multiple times between two weeks and two years of age. Thus both the study design and age range for the pre-PCV7 cohorts are different in comparison with the cohorts in the post-PCV7 and post-PCV13 periods which are cross-sectional in design and have age limits up to 4-5 years.

Prior to PCV7 introduction, PCV7 serotypes were identified on 9.8-15.2% of children's swabs (table 8). At the time of PCV7 introduction, from September 2006, PCV7 serotypes were present on 17.3% of children's swabs, with progressive declines seen in cohorts sampled in subsequent years. The most recent study showed PCV7 serotypes were only detected on 0.4% of children's swabs.

The additional six serotypes included in PCV13 were present on 2.2-3.3% of children's swabs before PCV7 introduction. Following PCV7 introduction an increase was seen with 4.6-9.9% of children's swabs having one of the additional PCV13 serotypes detected. The most recent studies following PCV13 introduction report only 0.4-0.9% of children swabbed having one of the additional PCV13 serotypes detected.

Non-vaccine type pneumococci were detected on 5.3-8.8% of children's swabs and consistently increased over the periods of PCV7 and then PCV13 introduction, with the most recent studies demonstrating 33.6-46.9% of children swabbed carry a NVT. Serotype-specific prevalence of NVTs are detailed in appendix A table 38.

Table 7. ***Characteristics of cohorts from studies reporting serotype-specific pneumococcal carriage in UK children.***

First author (publication year)	Sleeman (2006)*	Meats (2003)*	Tocheva (2011)	Tocheva (2011)	Tocheva (2011)	Flasche (2011)	Devine (2017)	Devine (2017)	Hamaluba (2015)	Devine (2017)	van Hoek (2014)	Devine (2017)
Time period of collection	Jul. 1999 - May 2001	Jul. 1999 - May 2001	Oct. 2006 - Mar. 2007	Oct. 2007 - Mar. 2008	Oct. 2008 - Mar. 2009	Apr. 2008 - Nov. 2009	Oct. 2009 - Mar. 2010	Oct. 2010 - Mar. 2011	Nov. 2010 - Sep. 2011	Oct. 2011 - Mar. 2012	Jul. 2012 - Mar. 2013	Oct. 2012 - Mar. 2013
PCV period	Pre-PCV7		Post-PCV7					Post-PCV13				
Age group	2 weeks - 2 years	2 weeks - 2 years	< 4 years	< 4 years	< 4 years	< 5 years	< 4 years	< 4 years	25-55 months	< 4 years	< 5 years	< 4 years
Study design	Longitudinal	Longitudinal	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional
Number of swabs analysed	1907	1353	324	373	328	192	399	287	575	333	277	223

Table 8. ***Serotype-specific pneumococcal carriage prevalence in UK children.***

Serotype	Sleeman (2006)*	Meats (2003)*	Tocheva (2011)	Tocheva (2011)	Tocheva (2011)	Flasche (2011)	Devine (2017)	Devine (2017)	Hamaluba (2015)	Devine (2017)	van Hoek (2014)	Devine (2017)
4	2 (0.1)	1 (0.07)	1 (0.31)	-	1 (0.3)	-	-	-	-	-	-	-
6B	108 (5.66)	44 (3.25)	24 (7.41)	15 (4.02)	4 (1.22)	2 (1.04)	2	1	2 (0.35)	-	-	-
9V	3 (0.16)	8 (0.59)	3 (0.93)	-	1 (0.3)	-	-	-	-	-	-	-
14	36 (1.89)	17 (1.26)	5 (1.54)	5 (1.34)	7 (2.13)	-	-	-	-	-	-	-
18C	23 (1.21)	7 (0.52)	-	-	-	2 (1.04)	1	-	-	-	-	-
19F	78 (4.09)	30 (2.22)	14 (4.32)	7 (1.88)	1 (0.3)	3 (1.56)	-	-	1 (0.17)	-	1 (0.36)	-
23F	40 (2.1)	26 (1.92)	9 (2.78)	10 (2.68)	3 (0.91)	1 (0.52)	3	-	1 (0.17)	-	-	-
1	-	-	-	-	1 (0.3)	-	1	1	4 (0.7)	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
7F	5 (0.26)	1 (0.07)	1 (0.31)	-	1 (0.3)	2 (1.04)	2	1	8 (1.39)	1	-	-
3	21 (1.1)	6 (0.44)	2 (0.62)	1 (0.27)	2 (0.61)	5 (2.6)	2	1	18 (3.13)	5	1 (0.36)	-
6A	29 (1.52)	14 (1.03)	9 (2.78)	16 (4.29)	6 (1.83)	2 (1.04)	5	2	-	1	-	2
19A	5 (0.26)	9 (0.67)	3 (0.93)	3 (0.8)	7 (2.13)	10 (5.21)	12	5	27 (4.7)	1	-	-
PCV7# %	15.2	9.8	17.3	9.9	5.2	4.2	1.5	0.3	0.7	-	0.4	-
Additional PCV13# %	3.2	2.2	4.6	5.4	5.2	9.9	5.5	3.5	9.9	2.4	0.4	0.9
NVT# %	8.8	5.3	9.3	11.5	20.1	37	20.8	31	36.7	27.3	46.9	33.6

Serotypes in bold and coloured green are those additional serotypes covered by PCV10 and PCV13. Serotypes in bold and coloured red are those additional serotypes covered by PCV13 only. *Sleeman 2006 and Meats 2003 both had a longitudinal study design. #Percentage of total number of swabs analysed. White columns = Pre-PCV7, green columns = Post-PCV7, and blue columns = Post-PCV13.

1.6.5 Serotype-specific distribution of pneumococcal carriage in Nepal: confirmation that PCV effect can be measured in resource-limited settings

Two carriage studies in Nepalese children have been published to date. Both studies assess carriage by community based children from Kathmandu (table 9). In addition, one of the studies also includes an assessment of carriage by community based children from the rural region of Okhaldhunga. This rural cohort had swabs collected and stored in silica desiccant packages (SDPs) until processing 2 weeks later and were compared with children from Kathmandu who had swabs collected and processed both following SDP storage for 2 weeks and after placement in skim-milk-tryptone-glucose-glycerin (STGG) with processing within 8 hours.

These studies consistently show that most of the serotypes covered by PCVs are in circulation, although serotypes 1, 5, and 7F, were infrequently identified across all of the study cohorts (table 10). Notably serotypes 14, 19F, and 23F were consistently the most frequently isolated PCV serotypes in each of the cohorts. Serotype-specific prevalence of NVTs for these cohorts are detailed in appendix A table 39.

Table 9. ***Characteristics of pneumococcal carriage studies conducted in Nepal.***

First author (publication year)	**Hanieh (2014)**	**Hanieh (2014)**	**Hanieh (2014)**	**Kandasamy (2015)**
Time period of collection	Oct. 2009 - Dec. 2009	Jan. 2009 - Sep. 2009	Jan. 2009 - Sep. 2009	May 2012 - Oct. 2012
Age group	6 weeks - 24 months	6 weeks - 24 months	6 weeks - 24 months	6 weeks - 24 months
Context	Healthy children	Healthy children	Healthy children	Healthy children
Region	Okhaldhunga	Kathmandu	Kathmandu	Kathmandu
Collection method	SDPs	SDPs	STGG	STGG
Serotyping method	PCR of single isolates	PCR of single isolates	PCR of single isolates	DNA microarray of plate sweeps
Swabs analysed, n	526	574	574	600
Children carrying pneumococcus, n (%)	364 (69.2)	235 (40.9)	337 (58.7)	309 (51.5)

SDPs=Silica desiccant packages.

Table 10. ***Serotype-specific pneumococcal carriage by Nepalese children prior to PCV10 introduction***

Serotype	Hanieh (2014) rural SDP, n (%)	Hanieh (2014) urban SDP, n (%)	Hanieh (2014) urban STGG, n (%)	Kandasamy (2015) urban STGG, n (%)
4	1 (0.2)	5 (0.9)	5 (0.9)	3 (0.5)
6B	1 (0.2)	2 (0.3)	1 (0.2)	12 (2)
9V[#]	-	-	-	7 (1.2)
14	11 (2.1)	10 (1.7)	17 (3)	19 (3.2)
18C	-	-	1 (0.2)	2 (0.3)
19F	8 (1.5)	10 (1.7)	10 (1.7)	18 (3)
23F	10 (1.9)	11 (1.9)	11 (1.9)	17 (2.8)
1	1 (0.2)	1 (0.2)	2 (0.3)	-
5	1 (0.2)	-	-	-
7F	-	1 (0.2)	-	-
3	1 (0.2)	2 (0.3)	6 (1)	10 (1.7)
6A	1 (0.2)	1 (0.2)	-	20 (3.3)
19A	4 (0.8)	2 (0.3)	5 (0.9)	6 (1)
PCV7 %	5.9	6.6	7.8	13
Additional PCV10 %	0.4	0.3	0.3	0
Additional PCV13 %	1.1	0.9	1.9	6
NVT %	61.8	33.1	48.6	32.5

[#]Serotypes 9A and 9V were not differentiated by PCR in Hanieh *et al* 2014. Three isolates from the urban SDP collection in Hanieh *et al* 2014 were identified as mixed serotypes.

1.7 Patterns of antibiotic use in the United Kingdom and Nepal

Pneumococcus is the most likely cause of bacterial pneumonia in children, and as such first-line antibiotic treatment in the community and hospitals needs to provide appropriate therapeutic cover. At least since the 1990's, amoxicillin has been recommended for use as first-line treatment of children with community acquired pneumonia in the UK.[106,107] Amoxicillin is the most frequently prescribed antibiotic in the UK with nearly 13 million prescriptions made in 2014.[108] This number of prescriptions appears to have been relatively static for at least the last decade.[108] In the UK macrolide antibiotics are recommended for use in the treatment of childhood pneumonia if there is no response to first-line therapy or if mycoplasma or chlamydia aetiologies are suspected.[107] Macrolides are second only to amoxicillin for the number of prescriptions dispensed each year in the UK. Notably the number of prescriptions have increased from just over 4 million in 2004 to nearly 5 million in 2014.

In Nepal general guidance on antibiotic use has historically been based around the guidance set out by the World Health Organization. Since 2010 treatment guidelines have been published by the Nepalese Ministry of Health and Population with the last update in 2014. Consistent with WHO recommendations up until 2014 the guidelines suggested the use of co-trimoxazole as first-line treatment of non-severe community acquired pneumonia.[109] Since 2014 however, WHO guidance has changed to having amoxicillin as first-line treatment.[110] Anecdotal evidence from paediatricians in Nepal indicates that this is the current guidance that is followed in practice.

There are no publically available data on antibiotic prescriptions made amongst Nepalese communities, and collection of these data would not necessarily reflect true antibiotic consumption, as up to 97% of street-based pharmacies have been shown to dispense antibiotics without a prescription.[111] Antibiotic use in Nepal is likely to broadly reflect that of India and China, which border Nepal and are its main trade partners. In 2014 is was estimated that in

India and China the four most consumed antibiotics by class were broad-spectrum penicillins, cephalosporins, macrolides, and fluoroquinolones.[112] Within these four classes of antibiotics, the greatest increase in use from 2004 to 2014 was seen for the cephalosporins.[112]

1.8 Patterns of pneumococcal antibiotic resistance in the United Kingdom and Nepal

Penicillin was first famously used in 1941 to treat an Oxfordshire policeman. However, it was just over two decades before penicillin resistant strains of pneumococcus were first reported in the 1960's.[113,114] Resistance to different antibiotic classes has since been observed as the new classes have been progressively introduced into clinical practice. It is now not uncommon in certain settings to detect strains which are resistant to multiple classes of antibiotics. The key factor influencing the prevalence of antibiotic resistance in a given setting is antibiotic usage.

In the United Kingdom the Government is now actively collecting data on penicillin resistance of pneumococcal disease isolates. The most recent data show that in 2014, 4.4% of invasive pneumococcal disease isolates tested were resistant to penicillin.[115]

In Nepal studies which have reported on penicillin resistance have all been conducted before the change in policy in 2014 to use amoxicillin as first-line treatment for pneumonia in children. Not surprisingly these studies show that penicillin resistance of pneumococci is relatively low, with reports ranging from 0-5% of isolates being resistant.[116–119] The recommendation for co-trimoxazole to be used as first-line antibiotic therapy up until 2014 is reflected in the contrastingly higher rates of resistance which range from 34-68%.[118,119]

1.9 Characterising pneumococcus using contemporary technologies.

For the first century following the discovery of pneumococcus, phenotypic characterisation, such as capsular serotyping and antibiotic resistance profiles, were developed and used for pneumococcal strain identification.

More recently, as new molecular technologies have been developed they have been progressively applied to the further characterisation of pneumococcus. The use of techniques such as restriction fragment length polymorphism, a process where genomic DNA is broken into fragments by restriction enzymes and the pattern of fragment size characterised, quickly led researchers to realise that pneumococcus undergoes horizontal gene transfer more frequently than many other bacterial species.[120] This also showed the importance of molecular characterisation of pneumococcus for providing insight into variation within phenotypic classifications.

1.9.1 Multi-locus sequence typing

Multiplex PCR allowed the targeted characterisation of pneumococci based on a set of seven specific housekeeping genes, known as multi-locus sequence typing (MLST). Pneumococci were then able to be classified into specific sequence types using this nomenclature. MLST thus became a key method for characterising pneumococcus at a population level, providing key insights into strain dissemination and emergence. The broad database of sequence types for pneumococci means that MLST continues to be a key resource in characterising pneumococcal populations from around the world.

1.9.2 Whole genome sequencing

The last two decades has seen the application of whole genome sequencing to pneumococcal science, with the first whole genome of pneumococcus published in 2001.[121] This and

subsequent whole genome sequences from other isolates demonstrate that the pneumococcal genome is haploid, approximately 2,200,000 bp in length, and contains about 2,200 coding regions.

The development of high-throughput sequencing platforms have made it feasible to sequence greater numbers of pneumococci and now provides a whole-genome depth of insight across a population. To date there are only a limited set of published studies which have conducted whole-genome sequencing of pneumococci across discrete population settings (table 11).

Croucher *et al* and Devine *et al* both examine isolates collected from resource rich settings and examine periods across the introduction of PCVs.[122,123] Devine *et al* primarily reports serotype data and does not present a molecular epidemiological analysis. Croucher *et al* showed that vaccine covered serotypes decreased in prevalence with time following vaccine introduction. However, many of the population clusters which contained vaccine serotypes showed expansion in non-vaccine types, indicating that the overall fitness of these strains had not been affected by the vaccine.

Chewapreecha *et al* and Everett *et al* each studied samples collected from resource-limited settings.[124,125] Everett demonstrates that those isolates with a serotype covered by PCV13 have a range of genetic diversities, for example serotype 1 had a very narrow range of genetic diversity compared with serotype 19F. The recognition of genetic diversity within strains being important as the findings by Croucher *et al* would suggest that strains with more genetic diversity have greater resilience to vaccine effects.

The study conducted by Chewapreecha *et al*, reports the largest number of pneumococcal isolates to have undergone whole genome sequencing in a pneumococcal population genomics study to date. This study had a longitudinal design and collected nasopharyngeal samples from infants and their mothers over a two-year period. This study focused on the dynamics of

recombination between pneumococcal strains, showing that non-typeable pneumococci act as the central exchange point for the transfer of DNA between strains.

It is important to note the limitations of those studies which have explored the population genomics of pneumococcus to date. Most notable of which are the lack of samples collected in the context of disease, the relatively narrow time-period over which samples have been collected, and the limited geographical regions from which the samples have been sourced. Broadening of the currently available data sets with collections from more regions around the world and across a greater expanse of time is needed if whole genome sequencing is to provide insight into fundamental knowledge gaps in pneumococcal biology. For example, the diversity of pneumococcus prior to widespread PCV use, the effect of antibiotic use on pneumococcal populations, and the differences between strains which predominantly colonise compared with strains which predominantly cause disease.

Table 11. ***Population based studies of pneumococci using whole genome sequencing***

First author (publication year)	Time period	Age group	Region	Setting	Vaccine period	Isolate source	Number of isolates sequenced
Croucher (2013)	2001-2007	Children aged 3 months to 7 years	Massachusetts, U.S.A	Healthy community based carriers	PCV7 introduction and post-PCV7	Carriage	616
Chewapreecha (2014)	2007-2010	Infants and their mothers	Mae La, Thailand	Healthy community based carriers	Pre-PCV	Carriage	3085
Devine (2017)	2006-2013	Children aged 4 years and under	Southampton, United Kingdom	Presentations to Southampton Hospital	Post-PCV7 and post-PCV13	Carriage	696
Everett (2012)	2003-2008	Children and adults	Blantyre, Malawi	Presentations to Queen Elizabeth Central Hospital	Pre-PCV13	IPD and carriage	127 (106 invasive and 21 carriage)

1.10 Aims and objectives of the thesis

This thesis aims to assess the effect PCV13 has had on pneumococcal carriage and disease in the UK. In addition, it aims to use pneumococcal samples collected from Nepalese children prior to PCV10 introduction; to evaluate the pneumococcal population at a serological and molecular level, to assess a molecular approach to pneumococcal diagnostics, and to evaluate antimicrobial resistance in this setting.

Specifically, this thesis will address the following objectives:

- To determine the effect of PCV13 on carriage, sero-prevalence, and IPD in UK children.
- To determine serotype-specific distribution of pneumococcal carriage in Nepalese children prior to vaccine introduction.
- To evaluate multi-serotype carriage in Nepalese children prior to PCV10 introduction.
- To evaluate the utility of carriage density in the diagnosis of pneumococcal pneumonia.
- To evaluate the molecular epidemiology of pneumococci isolated from Nepalese children prior to PCV10 introduction.
- To evaluate antibiotic resistance patterns in Nepal with a molecular focus.
- To determine the global population structure of pneumococci prior to PCV introduction.

Chapter 2 Persistent circulation of vaccine serotypes and serotype replacement with high attack-rate non-vaccine types after five years of UK infant immunisation with PCV13.

2.1 Introduction

Before introduction of the 7-valent pneumococcal conjugate vaccine (PCV7), *Streptococcus pneumoniae* was responsible for over 8000 cases of invasive pneumococcal disease (IPD) per year across all ages in England and Wales.[77] PCV7 administered at 2, 4, and 12-13 months, coupled with a catch-up programme for children aged up to 24 months, was introduced into the UK infant schedule in September 2006.[126] A reduction in carriage and IPD due to serotypes covered by PCV7 was initially observed in the vaccine-eligible cohort, followed by a decline in older, unvaccinated children and adults (herd protection).[77,103,127] At the same time, there was an increase in non-vaccine type (NVT) IPD (serotype replacement) across all age groups, particularly serotypes 1, 7F, and 19A.[103] In April 2010 the 13-valent PCV (PCV13), covering the additional serotypes 1, 3, 5, 6A, 7F, and 19A, replaced PCV7 in the UK infant schedule.

PCV13 introduction in the UK led to a reduction in overall PCV13-type disease and carriage, with just over 4000 IPD cases occurring across all ages in England and Wales between July 2013 and June 2014.[90,128] A similar effect on carriage was also seen with a study showing a reduction in children <5 years of age carrying a PCV13 serotype from 19.3% four years prior to PCV7 introduction to <1% two years after PCV13 introduction in the UK.[128] At the serotype level, the effect is less consistent, exemplified by a trial comparing carriage in children randomised to receive either PCV7 or PCV13, which described no detectable effect of PCV13 on serotype 3 carriage in Israeli children prior to community-wide PCV13 roll-out, and another report showing no significant reduction of serotype 3 IPD in UK children <5 years of age, 4 years after PCV13 introduction.[90,129,130] For serotype 19A, reductions in IPD following PCV13 introduction in the UK have been shown, however the magnitude of response is less than with

most other PCV13 types, with the number of cases in children <5 years returning to approximately those seen prior to PCV7 introduction (7.5 cases/year pre-PCV7 versus 7 cases/year post-PCV13), and the number of cases in adults 18-64 years still higher than the pre-PCV7 baseline (21.7 cases/year pre-PCV7 versus 104 cases/year post-PCV13).[77,90] More recent reports from Public Health England (PHE) indicate that serotype 19A disease is beginning to rise from the trough incidence seen in the year of PCV13 introduction.[131]

The rapid increase in serotype replacement following PCV7 introduction, in particular, NVTs which were infrequently identified among asymptomatic carriers but often cause disease (high attack-rate), raises the question as to whether a similar phenomenon would occur post-PCV13.[77,103] To date however, there has only been a statistically significant rise in serotype 24F IPD among English and Welsh children <5 years since PCV13 introduction.[90]

2.1.1 Study rationale

In this study, we describe PCV13 serotype-specific effects and emergence of serotype replacement through the examination of serotype-specific pneumococcal carriage, sero-prevalence of antibody against vaccine serotypes, and IPD in UK children. In addition, we describe the effects of infant vaccination with PCV13 on herd protection through the analysis of carriage and IPD in UK adults.

2.2 Methods

2.2.1 Study design

A cross-sectional observational study to establish the point-prevalence of pneumococcal nasopharyngeal carriage in children 13-48 months of age and their parents/legal guardians from the Thames Valley region of the United Kingdom was conducted between February 2014 and August 2015 (late-PCV13 era). Children were included in the study who were in good health and had received three doses of PCV13, at 2, 4, and 12 months of age, as per the recommended UK infant immunisation schedule. Children who had not received a complete course of PCV13, had taken antibiotics in the preceding 30 days, were febrile, had a respiratory illness, or had a health condition that may have influenced the study were excluded. Enrolled parents were in good health and resided in the same household as their participating child. Parents who were febrile, had a respiratory illness, or had a health condition that may influence the study were excluded. Demographic information about siblings, household smokers, day care attendance and antibiotic administration was recorded. All participants underwent sampling with a single nasopharyngeal swab. All participants had the option to also provide a blood sample. Data collected in this present study were compared with data from a similar study of children vaccinated with three doses of PCV7 conducted in the same region from November 2010-September 2011 (early-PCV13 era).[91]

2.2.2 Outcomes

The primary outcome was the presence of serotype 19A pneumococci on children's swabs. Secondary outcomes included, the presence of other pneumococcal serotypes on children's swabs; the presence of pneumococcal serotypes on parent's/legal guardian's swabs; the molecular serotype of nasopharyngeal carriage isolates from children and parents/legal guardians; the serotype-specific odds of carriage to invasive disease of isolates recovered from

children; and the serotype specific pneumococcal antibody levels in children and their parents/legal guardians.

2.2.3 Nasopharyngeal swab collection, microbiology, and microarray analysis

Swabs were collected and processed according to WHO guidelines.[132] A single flexible aluminium shaft with rayon tip (MWE, Wiltshire, UK) was passed through the anterior nares as far as the posterior pharynx, and then rotated 360 degrees before removal. The rayon swabs were cut into a tube of 0.5-1ml of skim-milk-tryptone-glucose-glycerine transport medium. Swabs were plated on Columbia-blood agar and incubated overnight before morphologically distinct alpha-haemolytic colonies were selected and sub-cultured overnight in the presence of optochin. Presumptive pneumococcal isolates underwent DNA extraction followed by molecular serotyping using the BUGS Senti-SP microarray. Secondary phenotyping by Quellung reaction (Serum Statens Institute, Denmark) was performed on variant genotypes.

2.2.4 Blood collection, processing, and serum measurement of IgG

Blood was collected from children using finger-prick. Following cleaning with alcohol and coating with petroleum jelly, the lateral border of the index finger was lanced and blood collected in a microtube (multivette 600, Sarstedt, Leister, UK). Blood was maintained between 4-8°C until processing. Pneumococcal serotype-specific serum IgG was measured by multiplex immunoassay (Bio-Rad Laboratories, Hercules, CA, USA) with Luminex technology.[133] Serum samples were analysed for IgG against the 13 serotypes covered by PCV13. The sera analysed were those from individuals who were carrying a PCV13 vaccine serotype (serotype 3, n=7; serotype 19A, n=8; and serotype 19F, n=1) and a random selection of NVT carriers (n=98), and non-carriers (n=107).

2.2.5 IPD surveillance

IPD data were obtained from the active surveillance of cases conducted across England and Wales by Public Health England (PHE). Pneumococci were serotyped by PHE using PCR. Data were extracted from this database for the two time periods between November 2010-September 2011 and February 2014-August 2015.

2.2.6 Statistical analysis and study size

To calculate the sample size for the primary outcome we used the prevalence of 4.7% for serotype 19A in PCV7 vaccinated children and determined that 600 participants would be required to produce a two-sided 95% confidence interval with a width equal to 3.6% (i.e. 95% CI = (3.1% to 6.7%)). A sample size of 600 would provide 90% power to detect an absolute reduction in prevalence of serotype 19A of 2.4% (i.e. from 4.7% to 2.3%) at 5% level of significance.

Therefore, to be able to compare the children recruited in this study with those in the previous study for the prevalence of 19A we aimed to recruit 600 children aged 21-48 months. We also recruited an extra 400 children aged 13-48 months to provide a cohort for future matching across the whole age range for future cross-sectional studies and to provide a point estimate of carriage among children. In the paediatric population the overall carriage rate prior to the introduction of PCV13 was approximately 47%.[91]

Demographic characteristics of carriers and non-carriers were compared using t-tests or Mann-Whitney U tests for continuous variables, and Fisher's Exact test for categorical variables. The odds of carriage of each serotype in the early-PCV13 and post-PCV13 populations were compared using logistic regression models, adjusted for age and sex, and results presented as serotype-specific odds ratios with associated 95% confidence intervals and p values. For serotypes with no instances of carriage for one of the time periods, Fisher's Exact test was used to compute p values.

To assess the relative invasiveness of each serotype, serotype-specific carriage was compared to IPD using odds-ratios.

Odds ratios of the proportion of invasive disease caused by each serotype between the two measured time periods were used to determine if there was a change in the odds of a given serotype being detected. 2x2 tables were constructed and Fisher's Exact test used to compute p values. A p value threshold of 0.01 was used to assess significance.

Serum antibody concentrations were log-transformed and geometric mean concentrations were compared between children carrying different serotypes by calculating the geometric mean ratio adjusted for age and sex using a linear regression model. Serotype-specific antibody for any child which was substantially higher than median antibody levels in children the same age, was considered to represent indirect evidence of recent colonisation and subsequent natural boosting of antibody. Children less than 2 years of age were grouped according to 6-month age categories and those with antibody higher than the 75^{th} percentile + 1.5 × the interquartile range for their age-group (i.e. points indicated as outliers in a box plot) were considered to have high antibody levels and to have likely experienced a recent colonisation event. For children older than 2 years of age, the threshold for 18 month-2 year old children was used as antibody levels plateaued during the third year of life.

All statistical analyses were performed using R version 3.3.1.[134]

2.3 Results

Following screening and informed consent 1005 children and 204 parents were enrolled into the study (figure 2). There were 17 children who withdrew resulting in 988 swabs for analysis. Blood samples from 220 children were tested for pneumococcal antibodies. From the parent cohort four participants withdrew resulting in 200 swabs for analysis. The majority of participating parents were female (table 12). During the carriage sample collection period there were 408 cases of IPD in children under 5 years of age and 2697 cases in 18-65 year olds across England and Wales.

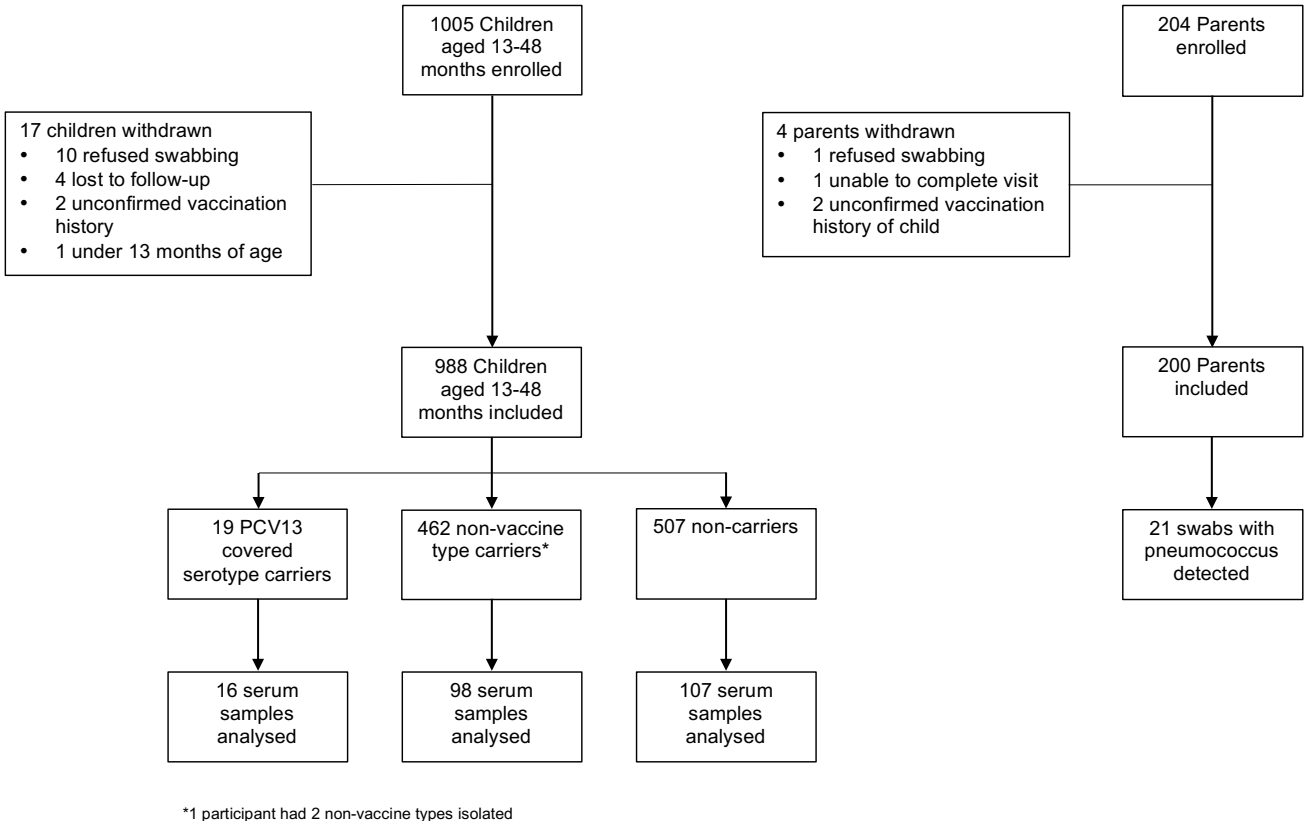


Figure 2. ***Overview of study recruitment.***
UK Children who had received three doses of PCV13 along with a subset of their parents were recruited into the carriage study between February 2014 and August 2015. A subset of serum samples collected from children at the same study visit, were analysed for IgG specific for each of the PCV13 covered serotypes. IPD data from enhanced national surveillance was extracted from the database held by the Department of Public Health England for the corresponding time-period.

Table 12. ***Characteristics of the child and adult carriage cohorts.***

Children		**All (N=988)**	**Carriers (N=481)**	**Non-carriers (N=507)**	**p***
Mean age, years (SD)		2.5 (0.9)	2.5 (0.9)	2.5 (0.9)	0.2007#
Males, N (%)		496 (50.2)	257 (53.3)	239 (47.1)	**0.0488**†
Ethnicity, N (%)	White	884 (89.5)	434 (90.2)	450 (88.8)	
	African Heritage /Afro-Caribbean	22 (2.2)	9 (1.9)	13 (2.6)	
	Asian	20 (2)	12 (2.5)	8 (1.6)	
	Other	62 (6.3)	26 (5.4)	36 (7.1)	
Household size	Mean (SD)	3.9 (0.9)	3.9 (0.9)	3.9 (0.9)	0.7547#
	Median [min, max]	4 [0, 8]	4 [2, 8]	4 [0, 7]	
Number of siblings	Mean (SD)	0.9 (0.8)	0.9 (0.9)	0.9 (0.8)	0.464§
	Median [min, max]	1 [0, 5]	1 [0, 5]	1 [0, 5]	
Children with a smoker in the house	n (%) [95% CI]	106 (10.7) [8.9-12.8]	45 (9.4) [6.9-12.3]	61 (12) [9.3-15.2]	0.155†
Daycare or preschool attendance per week	Mean hours (SD)	15.5(12.3)	16.2 (12.1)	14.9 (12.4)	0.0506§
	Median hours [min, max]	15 [0, 52]	15 [0, 50]	15 [0, 52]	
Adults		**All (N=200)**	**Carriers (N=21)**	**Non-carriers (N=179)**	**p***
Mean age, years (SD)		36.8 (5)	35.8 (5)	36.9 (5)	0.3448#
Males, N (%)		20 (10)	1 (4.8)	19 (10.6)	0.7011†
Ethnicity, N (%)	White	189 (94.5)	21 (100)	168 (93.9)	
	African Heritage /Afro-Caribbean	1 (0.5)	0 (0)	1 (0.6)	
	Asian	4 (2)	0 (0)	4 (2.2)	
	Other	6 (3)	0 (0)	6 (3.4)	
Household size	Mean (SD)	3.8 (0.9)	3.8 (0.8)	3.8 (1)	0.828§
	Median [min, max]	4 [2, 7]	4 [3, 5]	4 [2, 7]	
Parents with a smoker in the house	n (%) [95% CI]	23 (11.5) [7.4-16.8]	2 (9.5) [1.2-30.4]	21 (11.7) [7.4-17.4]	1†
Parents who smoke	n (%) [95% CI]	11 (5.5) [2.8-9.6]	1 (4.8) [0.1-23.8]	10 (5.6) [2.7-10]	1†
Parents with a child who was a carrier	n (%) [95% CI]	105 (52.5) [45.3-59.6]	12 (57.1) [34-78.2]	93 (52) [44.4-59.5]	0.8180†

2.3.1 Serotype-specific nasopharyngeal carriage in the post-PCV13 era

481/988 (48.7%, 95% CI 45.5-51.9%) children and 21/200 (10.5%, 95% CI 6.6-15.6%) adults were carrying at least one serotype of pneumococcus in their nasopharynx (table 12). One swab from a child had two serotypes (17F and NT) detected following culture, whilst five samples from children had greater than one *cps* locus detected by the microarray and were classified as mixed (figure 3). Carriers and non-carriers were similar in age, household makeup, and prevalence of smoking. Male children were significantly more likely to be carriers than females (p=0.0488) (table 12). There was no significant difference in the weekly hours spent at daycare and/or preschool between children who were identified as carriers compared with non-carriers (p=0.0506).

In the full child cohort, non-vaccine types (NVTs) represented the majority of carriage isolates (375/482, 77.8%), followed by vaccine related types (VRTs) (88/482, 18.3%), the additional serotypes in PCV13 but not PCV7 (18/482, 3.7%), and PCV7 only serotypes (1/482, 0.2%) (figure 3). Serotypes 19A and 3, which are additional serotypes covered by PCV13 but not PCV7, were both identified on 9/482 (1.87%) occasions each. Serotype 19F (covered by PCV7 and PCV13) was identified once. Serotype 15B was the most prevalent serotype, identified on 59/482 (12.2%) occasions. Serotype 3 carriers were significantly older (mean age 3.1 years) than NVT carriers (mean age 2.5 years) (p=0.0446) whilst there were no detectable differences in age between serotype 19A (mean age 2.1 years) and NVT carriers (p=0.087) (appendix B figure 39).

The effect on transmission and hence herd protection from carriage was examined via the parent cohort, within which there were no detected vaccine serotypes from the 21 pneumococcus positive swabs (figure 3). The most prevalent serotypes identified were 21, 23B, and 38 (3 occasions each).

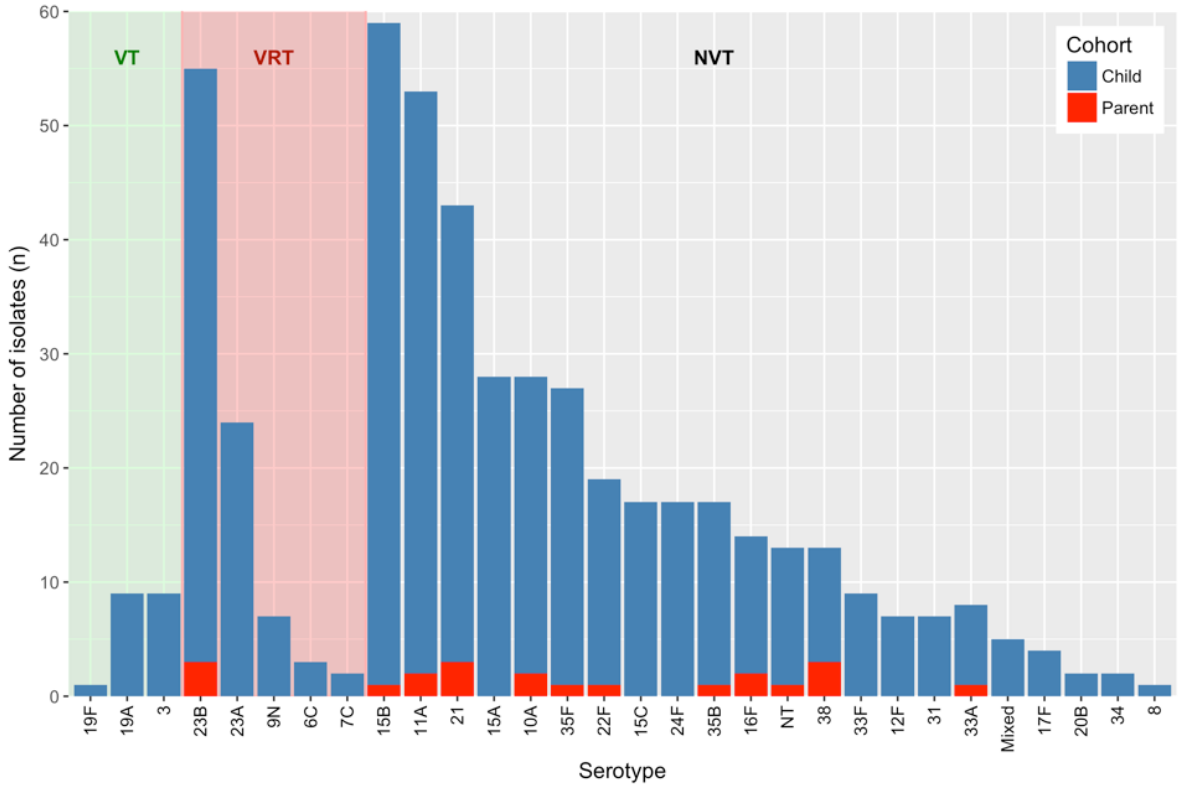


Figure 3. ***Serotype-specific prevalence and relatedness of pneumococcal isolates detected by DNA microarray of isolates from nasopharyngeal swabs collected from UK children who had received three doses of PCV13 and their parents between February 2014 and August 2015.***
481/988 children and 21/200 parents had at least one serotype of pneumococcus isolated from their respective nasopharyngeal swabs. VT = vaccine (PCV13) serotypes. VRT = vaccine (PCV13) related serotypes (that is serotypes within the same sero*group* as PCV13 serotypes). NVT = non-vaccine types.

2.3.2 Serotype-specific comparison of pneumococcal types identified in carriage with IPD samples

During the study period, in children under 5 years of age, the most common IPD serotypes isolated were 12F (49/408, 12%), 24F (30/408, 7.4%), and 23B (28/408, 6.9%). 11 serotypes (5.4% of isolates), were detected in IPD but not in carriage of which serotype 27 was most common (5/408, 1.2%). PCV13 covered serotypes accounted for 11.8% (48/408) of IPD cases. The odds of each serotype being detected in IPD samples compared with carriage samples, was higher in children for non-vaccine covered serotypes 8 and 12F (table 13).

The extent of herd protection in the unvaccinated population was also examined through the analysis of serotype-specific IPD in adults aged 18-65 years from England and Wales. In this age group 554/2697 (20.5%) of IPD cases were serotypes covered by PCV13. Serotypes 8 (544/2697, 20.2%) and 12F (391/2697, 14.5%) were the most common IPD-causing isolates within this cohort.

Table 13. ***Serotype-specific case-carrier ratios determined by comparing IPD in children <5 years from England and Wales five years after the introduction of PCV13 with carriage in children aged 13-48 months from the Thames Valley vaccinated with PCV13.***

Serotype	IPD cases[†], n (%), N=408	Carriage cases[*], n (%), N=482	OR (95% CI)	p
Serotypes covered by PCV7				
19F	8 (1.96)	1 (0.21)	9.62 (1.2-77.24)	0.014
Additional serotypes covered by PCV13				
1	1 (0.25)	-	-	0.4584
3	18 (4.41)	9 (1.87)	2.43 (1.08-5.46)	0.0312
5	2 (0.49)	-	-	0.2099
7F	4 (0.98)	-	-	0.0438
19A	15 (3.68)	9 (1.87)	2.01 (0.87-4.63)	0.1019
Vaccine related serotypes				
6C	5 (1.23)	3 (0.62)	1.98 (0.47-8.34)	0.4807
7C	-	2 (0.41)	-	0.5029
9L	1 (0.25)	-	-	0.4584
9N	15 (3.68)	7 (1.45)	2.59 (1.05-6.42)	0.049
18A	1 (0.25)	-	-	0.4584
18F	1 (0.25)	-	-	0.4584
23A	2 (0.49)	24 (4.98)	0.09 (0.02-0.4)	**<0.0001**
23B	28 (6.86)	52 (10.79)	0.61 (0.38-0.98)	0.0456
Non-vaccine types				
8	21 (5.15)	1 (0.21)	26.1 (3.5-194.9)	**<0.0001**
10A	24 (5.88)	26 (5.39)	1.1 (0.62-1.94)	0.7719
11A	8 (1.96)	51 (10.58)	0.17 (0.08-0.36)	**<0.0001**
12F	49 (12.01)	7 (1.45)	9.26 (4.15-20.69)	**<0.0001**
15A	-	28 (5.81)	-	**<0.0001**
15B/C[#]	27 (11.52)	75 (15.56)	0.39 (0.24-0.61)	**<0.0001**
16F	2 (0.49)	12 (2.49)	0.19 (0.04-0.87)	0.0269
17F	3 (0.74)	4 (0.83)	0.89 (0.2-3.98)	1
20	3 (0.74)	-	-	0.096
20B	-	2 (0.41)	-	0.5029
21	7 (1.72)	40 (8.3)	0.19 (0.09-0.44)	**<0.0001**
22F	24 (5.88)	18 (3.73)	1.61 (0.86-3.01)	0.1538
24B	2 (0.49)	-	-	0.2099
24F	30 (7.35)	17 (3.53)	2.17 (1.18-4)	0.0153
27	5 (1.23)	-	-	0.02
29	1 (0.25)	-	-	0.4584
31	1 (0.25)	7 (1.45)	0.17 (0.02-1.36)	0.0767
33A	-	7 (1.45)	-	0.0175
33F	20 (4.9)	9 (1.87)	2.71 (1.22-6.02)	0.0132
34	-	2 (0.41)	-	0.5029
35A	1 (0.25)	-	-	0.4584
35B	9 (2.21)	16 (3.32)	0.66 (0.29-1.5)	0.4162
35F	8 (1.96)	26 (5.39)	0.35 (0.16-0.78)	0.0081
38	13 (3.19)	10 (2.07)	1.55 (0.67-3.58)	0.397
NT	-	12 (2.49)	-	**0.0007**

[*]Five carriage samples were mixed. [†]29 IPD isolates were not serotyped. NT=non-typeable. [#]The final serotype of some disease causing 15B/C pneumococci could not be determined. P values <0.01 are in bold font.

2.3.3 Effect of PCV13 on pneumococcal carriage and IPD

A prior study examining the incidence of pneumococcal carriage in the Thames Valley region in the initial year of PCV13 introduction was compared with the current study for those participants within the overlapping age range of 24-48 months (N=268/567 (47.2%) early-PCV13 carriers, vs N=305/642 (47.5%) post-PCV13 carriers).[91] A significant reduction in carriage post-PCV13 were seen for the PCV13 serotype 19A (table 14). Serotype 6C also became less common. There was a significant increase in carriage due to serotypes 15A and 15B. The PCV13 serotypes 1, 6B, 19F, and 23F were detected in small numbers in the early-PCV13 cohort but were not detected in the late-PCV13 cohort.

408 IPD cases in 18 months occurred in children <5 years during the late PCV13 era, compared with 98 cases in total for the 10-month observation period in the year of PCV13 introduction. In children <5 years of age, there was a lower proportion of IPD after the introduction of PCV13 due to serotypes 1, 7F, and 19A, but no change in proportion of IPD due to serotypes 3, 5. No cases of 6A occurred during either period (table 15). In contrast there were significantly higher proportions of IPD due to serotype 12F and, 24F in the post-PCV13 compared with the early-PCV13 cohort. A rise in the proportion of IPD due to serotype 8 (0/98 to 21/408 cases) was seen although the respective p value was not below the 0.01 threshold set. In adults aged 18-65 years, 2697 IPD cases over 18 months were observed during the late PCV13 era compared with 189 during the 10-month observation period at the time of vaccine introduction. There was a significant decrease in the proportion of IPD for PCV7 serotypes 6B, 9V, and 23F from late-PCV13 compared with early-PCV13 adults. There was also a significant decrease in the proportion of IPD caused by the additional PCV13 covered serotypes 1, 3, 6A, 7F, and 19A; for vaccine related serotype 6C; and for three other non-vaccine serotypes (11A, 16F, 22F, and 38). Increases were observed for serotypes 8, 9N, 10A, and 12F.

Table 14. ***Serotype-specific carriage in PCV7 compared with PCV13 vaccinated children aged 24-48 months from the Thames Valley.***

Serotype	PCV7 vaccinated (early-PCV13) n (%) N=567	PCV13 vaccinated (late-PCV13) n (%) N=642	OR (95% CI) PCV7 vaccinated/ PCV13 vaccinated	p
Serotypes covered by PCV7				
6B	2 (0.35)	-	-	0.2197*
19F	1 (0.18)	-	-	0.4690*
23F	1 (0.18)	-	-	0.4690*
Additional serotypes covered by PCV13				
1	1 (0.18)	-	-	0.4690*
3	18 (3.17)	7 (1.09)	0.36 (0.14-0.83)	0.0224
7F	8 (2.99)	-	-	0.0022*
19A	27 (4.76)	3 (0.47)	0.08 (0.02-0.25)	**<0.0001**
Vaccine-related serotypes				
6C	24 (4.23)	3 (0.47)	0.11 (0.03-0.32)	**0.0004**
7C	-	1 (0.16)	-	1*
9N	-	6 (1.97)	-	0.0324*
23A	12 (2.12)	17 (2.65)	1.19 (0.55-2.63)	0.6543
23B	24 (4.23)	33 (5.14)	1.29 (0.75-2.25)	0.356
Non-vaccine types				
8	-	1 (0.16)	-	1*
10A	10 (1.76)	21 (3.27)	1.93 (0.91-4.34)	0.0959
11A	24 (4.23)	37 (5.76)	1.28 (0.74-2.22)	0.3807
12F	2 (0.35)	3 (0.47)	1.45 (0.24-11.1)	0.684
15A	4 (0.71)	19 (2.96)	4.37 (1.61-15.24)	**0.0081**
15B	2 (0.35)	32 (4.98)	13.85 (4.12-86.17)	**0.0004**
15C	8 (1.41)	9 (1.4)	0.92 (0.33-2.54)	0.8714
16F	8 (1.41)	5 (0.78)	0.37 (0.09-1.26)	0.1264
17F	1 (0.18)	2 (0.31)	1.32 (0.09-31.05)	0.834
20B	-	1 (0.16)	-	1*
21	12 (2.12)	27 (4.21)	1.98 (1-4.15)	0.0565
22F	13 (2.29)	12 (1.87)	0.86 (0.38-1.91)	0.7029
24F	1 (0.18)	12 (1.87)	10.87 (2.1-199.08)	0.0227
29	5 (1.87)	-	-	0.0219*
31	7 (1.23)	6 (0.93)	0.82 (0.26-2.5)	0.731
33A	-	2 (0.31)	-	0.5015*
33F	7 (1.23)	5 (0.78)	0.52 (0.14-1.76)	0.304
34	-	2 (0.31)	-	0.5015*
35B	11 (1.94)	11 (1.71)	0.94 (0.4-2.22)	0.8872
35F	15 (2.65)	11 (1.71)	0.55 (0.23-1.26)	0.1675
38	4 (0.71)	8 (1.25)	1.8 (0.55-6.86)	0.3451
NT	9 (1.59)	9 (1.4)	0.96 (0.37-2.48)	0.928

Odds ratios (OR) and the associated p values were adjusted for age and sex unless otherwise specified. P values <0.01 are in bold font. PCV7=seven-valent pneumococcal conjugate vaccine. PCV13=thirteen-valent pneumococcal conjugate vaccine. NT=non-typeable. *Fisher's Exact test. Time-period of collection from PCV7 vaccinated children; November 2010-September 2011. Time-period of collection from PCV13 vaccinated children; February 2014-August 2015.

Table 15. ***Serotype-specific proportions of IPD in children and adults in England and Wales, before and after the introduction of PCV13.***

	Children <5 years				Adults 18-64 years			
Serotype	**Cases (%) early-PCV13*** N=98 cases in 10 months	**Cases (%) late-PCV13*** N=408 cases in 18 months	**OR (95% CI)†**	**p**	**Cases (%) early-PCV13*** N=1892 cases in 10 months	**Cases (%) late-PCV13*** N=2697 cases in 18 months	**OR (95% CI)†**	**p**
Serotypes covered by PCV7								
4	-	-	-	-	10/1892 (0.53)	16/2697 (0.59)	1.12 (0.51-2.48)	0.8437
6B	-	-	-	-	31/1892 (1.64)	5/2697 (0.19)	0.11 (0.04-0.29)	**<0.0001**
9V	-	-	-	-	22/1892 (1.16)	4/2697 (0.15)	0.13 (0.04-0.37)	**<0.0001**
14	-	-	-	-	17/1892 (0.9)	11/2697 (0.41)	0.45 (0.21-0.97)	0.0521
18C	-	-	-		11/1892 (0.58)	10/2697 (0.37)	0.64 (0.27-1.5)	0.3748
19F	1/98 (1.02)	8/408 (1.96)	1.9 (0.24-15.7)	1	26/1892 (1.37)	26/2697 (0.96)	0.7 (0.4-1.21)	0.2046
23F	-	-	-	-	22/1892 (1.16)	4/2697 (0.15)	0.13 (0.04-0.37)	**<0.0001**
Additional serotypes covered by PCV13								
1	23/98 (23.47)	1/408 (0.25)	0 (0-0.06)	**<0.0001**	63/1892 (3.33)	49/2697 (1.82)	0.54 (0.37-0.78)	**0.0013**
3	5 (5.1)	18/408 (4.41)	0.9 (0.31-2.37)	0.7872	207/1892 (10.94)	135/2697 (5.01)	0.43 (0.34-0.54)	**<0.0001**
5	-	2/408 (0.49)	-	1	2 (0.11)	2/2697 (0.07)	0.7 (0.1-4.98)	1
6A	-	-	-	-	44/1892 (2.33)	5/2697 (0.19)	0.08 (0.03-0.2)	**<0.0001**
7F	20/98 (20.41)	4/408 (0.98)	0 (0.01-0.12)	**<0.0001**	213/1892 (11.26)	153/2697 (5.67)	0.47 (0.38-0.59)	**<0.0001**
19A	16/98 (16.33)	15/408 (3.68)	0.2 (0.09-0.41)	**<0.0001**	267/1892 (14.11)	134/2697 (4.97)	0.32 (0.26-0.4)	**<0.0001**
Vaccine-related serotypes								
6C	-	5/408 (1.23)	-	0.5886	89/1892 (4.7)	36/2697 (1.33)	0.27 (0.19-0.41)	**<0.0001**
7C	-	-	-	-	-	3/2697 (0.11)	-	0.2729
9N	-	15/408 (3.68)	-	0.0881	40/1892 (2.11)	144/2697 (5.34)	2.61 (1.83-3.73)	**<0.0001**
23A	-	2/408 (0.49)	-	1	54/1892 (2.85)	51/2697 (1.89)	0.66 (0.45-0.97)	0.0351
23B	2/98 (2.04)	28/408 (6.86)	3.5 (0.83-15.11)	0.0926	28/1892 (1.48)	50/2697 (1.85)	1.26 (0.79-2)	0.3555
Non-vaccine types								
8	-	21/408 (5.15)	-	0.0199	130/1892 (6.87)	544/2697 (20.17)	3.42 (2.8-4.19)	**<0.0001**
10A	3/98 (3.06)	24/408 (5.88)	2 (0.58-6.71)	0.3262	19/1892 (1)	75/2697 (2.78)	2.82 (1.7-4.68)	**<0.0001**
10F	-	-	-	-	3/1892 (0.16)	3/2697 (0.11)	0.7 (0.14-3.48)	0.6956
11A	1/98 (1.02)	8/408 (1.96)	1.9 (0.24-15.7)	1	38/1892 (2.01)	28/2697 (1.04)	0.51 (0.31-0.84)	**0.0079**
12B	-	-	-	-	1/1892 (0.05)	4/2697 (0.15)	2.81 (0.31-25.15)	0.6547
12F	2/98 (2.04)	49/408 (12.01)	6.6 (1.57-27.42)	**0.0014**	39/1892 (2.06)	391/2697 (14.5)	8.06 (5.76-11.26)	**<0.0001**
15A	3/98 (3.06)	20/408 (4.9)	1.6 (0.48-5.61)	0.5924	58/1892 (3.07)	90/2697 (3.34)	1.09 (0.78-1.53)	0.6714

15B/C[#]	4/98 (4.08)	27/408 (6.62)	0.6 (0.21-1.76)	0.4824	23/1892 (1.22)	48/2697 (1.78)	1.47 (0.89-2.43)	0.145
16F	-	2/408 (0.49)	-	1	36/1892 (1.9)	34/2697 (1.26)	0.66 (0.41-1.06)	0.0872
17F	2/98 (2.04)	3/408 (0.74)	0.4 (0.06-2.16)	0.2494	20/1892 (1.06)	23/2697 (0.85)	0.81 (0.44-1.47)	0.5345
20	-	3/408 (0.74)	-	1	21/1892 (1.11)	24/2697 (0.89)	0.8 (0.44-1.44)	0.4519
21	1/98 (1.02)	7/408 (1.72)	1.7 (0.21-13.92)	1	1/1892 (0.05)	2/2697 (0.07)	1.4 (0.13-15.49)	1
22F	6/98 (6.12)	24/408 (5.88)	1 (0.38-2.41)	1	167/1892 (8.83)	156/2697 (5.78)	0.63 (0.51-0.8)	**<0.0001**
24B	-	2/408 (0.49)	-	1	-	-	-	-
24F	-	30/408 (7.35)	-	**0.0029**	21/1892 (1.11)	49/2697 (1.82)	1.65 (0.99-2.76)	0.0658
27	-	5/408 (1.23)	-	0.5886	1/1892 (0.05)	-	-	0.4123
29	1/98 (1.02)	1/408 (0.25)	0.2 (0.01-3.84)	0.3502	1/1892 (0.05)	-	-	0.4123
31	2/98 (2.04)	1/408 (0.25)	0.1 (0.01-1.31)	0.0974	20/1892 (1.06)	19/2697 (0.7)	0.66 (0.35-1.25)	0.2525
33A	-	-	-	-	-	2/2697 (0.07)	-	0.5153
33F	4/98 (4.08)	20/408 (4.9)	1.2 (0.4-3.63)	1	57/1892 (3.01)	86/2697 (3.19)	1.06 (0.75-1.49)	0.7959
34	-	-	-	-	3/1892 (0.16)	-	-	0.07
35A	-	1/408 (0.25)	-	1	-	5/2697 (0.19)	-	0.0819
35B	-	9/408 (2.21)	-	0.2173	23/1892 (1.22)	24/2697 (0.89)	0.73 (0.41-1.3)	0.2992
35C	-	-	-	-	2/1892 (0.11)	-	-	0.1699
35F	1/98 (1.02)	8/408 (1.96)	1.9 (0.24-15.7)	1	26/1892 (1.37)	35/2697 (1.3)	0.94 (0.57-1.57)	0.896
38	1/98 (1.02)	13/408 (3.19)	3.2 (0.41-24.7)	0.3238	24/1892 (1.27)	14/2697 (0.52)	0.41 (0.21-0.79)	**0.0075**
37	-	-	-	-	1/1892 (0.05)	3/2697 (0.11)	2.11 (0.22-20.26)	0.6476

Serotypes 18A, and 18F were only found on one occasion in the early-PCV13 period in children. Serotypes 9L were only found on one occasion in the late-PCV13 period in children. Serotypes 7B, 11B, and 12A were only found on one occasion in the early-PCV13 period in adults. Serotypes 9L, 10B, 11C, and 15F were only found on one occasion in the post-PCV13 period in adults. 8 isolates were classified as 16A/F in early-PCV13 adults. P values <0.01 are in bold font. PCV13=Thirteen-valent pneumococcal conjugate vaccine. [#]The final serotype of some serotype 15B/C pneumococci could not be determined. [†]No OR could be calculated where there were no detected cases in either the early-PCV13 time-period (November 2010-September 2011) or late-PCV13 time-period (February 2014-August 2015). In the late-PCV13 period 29 and 196 isolates are yet to be serotyped in the child and adult age groups respectively.

2.3.4 Serotype-specific comparison of carriage to IgG levels in children

The seven children carrying serotype 3 who had serum assayed, had higher geometric mean IgG concentrations against serotype 3 than carriers of NVTs (GMR=5.98, 95% CI 1.86-19.16, p=0.003) or non-carriers (GMR=6.17, 95% CI 1.96-19.47, p=0.0022) (appendix B table 40).

The eight children carrying serotype 19A who had serum assayed, had higher IgG concentrations against serotype 19A and 19F than carriers of NVT (GMR=6.89, 95% CI 2.62-18.13, p=0.0001, and GMR=3.76, 95% CI 1.51-9.31, p=0.0048 for 19A and 19F respectively). Similar results were seen when comparing with non-carriers (appendix B table 41).

The geometric mean ratio (GMR) of IgG to serotypes 1 (GMR=0.44, 95% CI 0.22-0.87, p=0.0202), and 4 (GMR=0.44, 95% CI 0.21-0.95, p=0.0388), were significantly lower in serotype 3 carriers compared with NVTs carriers. The GMR of IgG to serotype 19F (GMR=0.31, 0.12-0.84, p=0.0226) was also significantly lower in serotype 3 carriers compared with non-carriers.

Raised antibody levels due to natural boosting occurred most frequently for serotypes 3 and 19A, with 24/204 (11.76%) and 14/204 (6.86%) of children having high antibody levels for these serotypes (figure 4, appendix B figure 40, and appendix B figure 41). Of these, 4/29 (13.79%) and 2/17 (11.76%) were also carrying serotype 3 or 19A/F respectively. 23 participants had high antibody levels for more than one PCV13 serotype with the maximum number being four. The mean age of children with naturally-boosted antibody levels was significantly higher than the average age of children sampled (3 vs 2.1 years, $p<0.0001$).

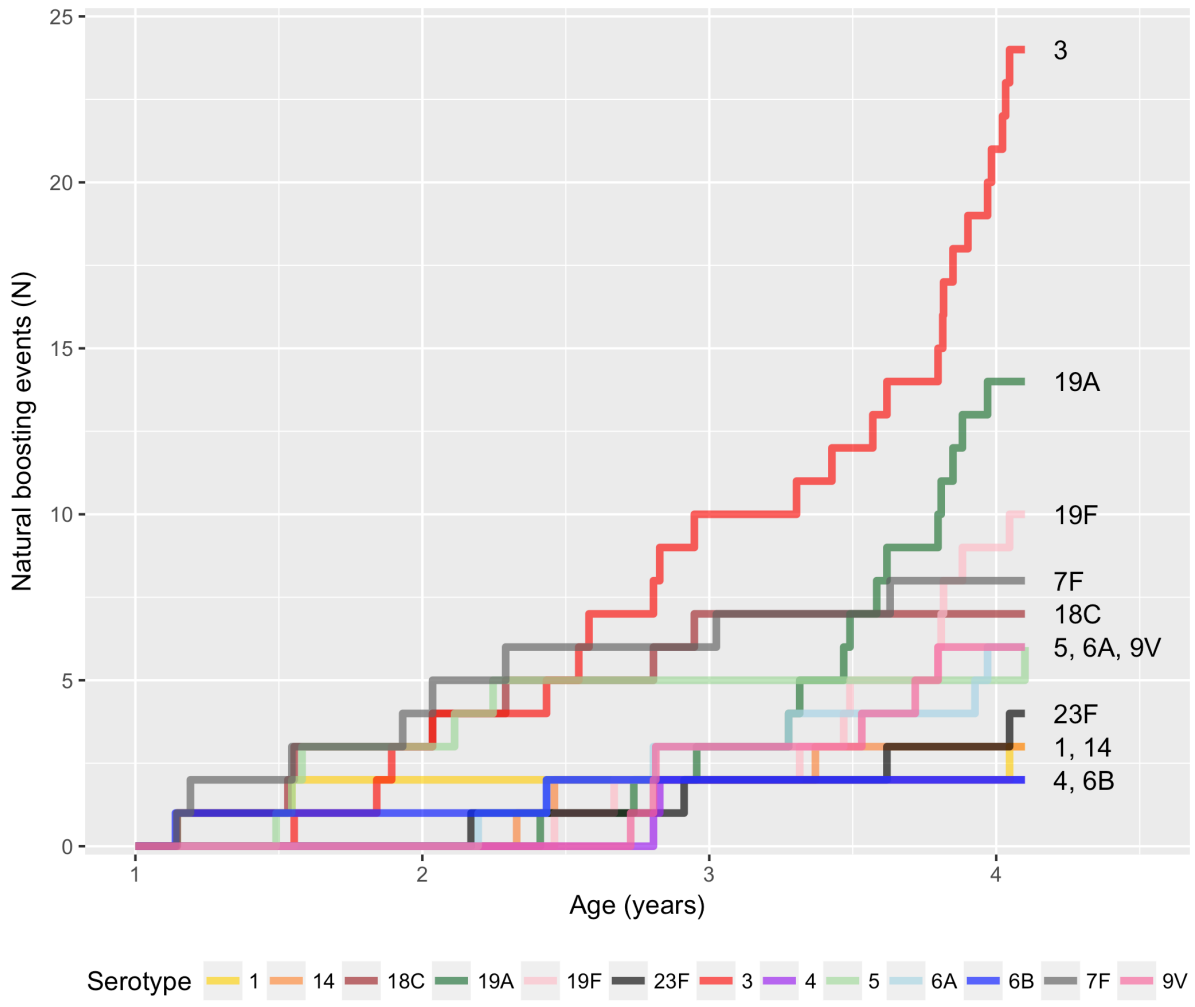


Figure 4. ***Cumulative number of natural boosting events according to age of UK children that had received 3 doses of PCV13.***
Serum collected by finger-prick from 204 PCV13 immunised children at the same visit as nasopharyngeal sampling for pneumococcal carriage and was analysed by multiplex immunoassay for antibodies to the 13 capsular polysaccharides contained in PCV13. Naturally boosted children were defined as those having a serotype-specific serum IgG level which was raised above the defined threshold. We defined the threshold as 1.5 times greater than the upper IQR and applied it across age-quartiles until there was an increase in the threshold, at which point the threshold was kept at the lowest limit it reached before it increased. All children included in this analysis were a random selection of either NVT carriers or non-carriers.

2.4 Discussion

Despite the apparently limited impact of PCV13 on serotype 19A, this study provides evidence for a reduction in carriage by PCV13 vaccinated children when compared with PCV7 vaccinated children. With further evidence provided by the nation-wide IPD surveillance amongst children under 5 years which shows a concurrent decrease when comparing the same time-periods. For serotype 3 there was no significant reduction in either carriage or IPD detected amongst children when comparing the two time periods. The fact that serotypes 3 and 19A, although reduced, were still the most frequently identified PCV13 serotypes carried by children suggests that PCV13 may have differential effects on each serotype. To explore this effect, we examined children's antibody levels against PCV13 serotypes and identified participants with higher than normal antibody levels for their age, which would be consistent with prior colonisation by circulating strains. These measurements were highest for pneumococcal serotype 3 followed by serotype 19A. Indicating that these serotypes are still frequently transmitted amongst children and hence might explain why a reduction in the proportion of childhood serotype 3 IPD was not detected. These findings provide evidence that support the use of PCV13 schedules which give higher antibody levels beyond the second year of life as they might provide better protection against vaccine type transmission.

2.4.1 Pre-PCV13 carriage and IPD compared with post-PCV13 carriage and IPD

Serotype replacement, via the expansion of VRTs and NVTs, appears to be responsible for overall carriage prevalence being unchanged between the two cohorts presented in this study, and is consistent with findings in other settings.[135] One interesting exception being the significant reduction in serotype 6C carriage in children, which may be due to cross-protective antibody generated towards closely related vaccine types 6A and 6B.[136]

Comparing PCV13 with PCV7 vaccinated children, there was an overall reduction in the proportion of vaccine types carried, with sample size numbers sufficiently large to demonstrate a significant reduction in serotype 19A. There was no significant reduction in serotype 3 carriage which is in keeping with a prior RCT in Israel showing no significant reduction in serotype 3 carriage in children receiving PCV13 compared with those who received PCV7.[129] It is important to note that in this current study however, that serotype 3 carriage had a strong trend towards a decrease between the two time frames, and had a larger sample size been used a significant change may have been detected.

There were few cases of serotype 3 IPD in children at the time of PCV13 introduction in the UK so it is difficult to demonstrate a decrease post-PCV13 from this low baseline. It is interesting to note however, that in children under 5 years in England and Wales, serotype 3 cases have progressed from 2.83 cases per year in 2000-2006 prior to PCV7, to 13 per year in 2008-2010 following PCV7, to 6 per year in 2010-2011 at the time of PCV13 introduction, to 8 per year in 2013-2014 and 12 per year in 2014-2015 following PCV13.[77,90] This indicates that serotype 3 IPD increased following PCV7 introduction and that the current number of cases of serotype 3 disease is approaching that seen in the post-PCV7 pre-PCV13 era. It is also important to note that the decrease in serotype 3 cases from 2008-2010 to 2010-2011 occurred very quickly after PCV13 introduction and could at least partly be due to natural cycles of disease outbreaks rather than vaccine effect.

In contrast to the overall reductions in IPD due to PCV13 covered serotypes, the significant increase in the proportion of IPD in children caused by serotypes 12F and 24F is concerning, with these two serotypes now accounting for 19% of IPD in children under 5 years, and 16% of adult cases. These serotypes have high attack-rates indicating that these strains colonise briefly, circulate rapidly, and possess features conferring high virulence. Increases in IPD due to serotypes 8, 10A, 12F, 15A, and 24F in English and Welsh residents aged 5-64 years have

previously been shown. However only an increase in serotype 24F was noted in children <5 years at that time.[90] Priority should be given to the inclusion of serotypes 12F and 24F in a higher-valence PCV as vaccination against these would have the greatest impact on disease, without substantially affecting the mix of less virulent serotypes found in asymptomatic carriage.

2.4.2 Detection of natural boosting events using serum antibody levels

The high number of natural boosting events elicited by serotypes 3 and 19A are reflective of the continued circulation of these serotypes throughout the community. The detection of serotype 3 carriage in healthy vaccinated children who are older than the other children in the study and have significantly lower serum immunoglobulin G levels against serotypes 1 and 4 than NVT carriers, support the conclusion that age-related decay in antibody levels creates susceptibility to carriage of serotype 3. The same pattern of findings is not as apparent for serotype 19A carriers with a number of carriers detected within the first year of boosting. This brings into question whether vaccine stimulated antibody generated to serotype 19A capsule polysaccharide needs to have higher concentrations and/or improved avidity to achieve the same effect on carriage as seen for other vaccine serotypes.

2.4.3 Strengths and limitations of the study

Limitations of the study are the cross-sectional design, female predominance in the parent carriage group, and differences between the carriage and IPD populations. The strong female predominance in the parent group needs to be considered since there was a relationship between sex and carriage in the child group. When making comparisons across numerous serotypes it is important to consider the possibility that some observations may be due to chance alone. Additionally, the reason why some observations may be apparent in the carriage data set, and

not reflected in the IPD data set is that the carriage population is a small subset of the national population. Thus there may be variations in one population that are not captured in the other, for example if an outbreak of disease due to a specific serotype occurred only in a discrete region of the nation a corresponding change in carriage may not be seen in the region where carriage samples were collected in this study.

Strengths of the study include the use of microarray technology to accurately detect serotypes, the sample size, and the time-point and timescale of collection. The microarray is able to discern closely related *Streptococcus spp.* which are often detected as NTs via Quellung. Seasonality has been shown to effect the prevalence of carriage, however the timescale of sample collection in this study spanned across all seasons reducing the likelihood of confounding due to this factor.

2.4.4 Conclusions

This study demonstrates a significant impact of PCV13 on carriage prevalence in children for serotype 19A but not serotype 3. There is also a significant reduction in carriage of the VRT 6C, an observation that supports cross-reactivity of serotype 6C with antibodies formed to serotypes 6A and 6B. These findings are coupled with significant reductions in the proportion of childhood IPD due to serotypes 1, 7F, and 19A but not serotype 3. Herd protection is evident by the absence of PCV13 covered serotypes carried by adults and reduction in the proportion of IPD due to serotypes 1, 3, 6A, 6B, 7F, 9V, 14, 19A, and 23F. We have also shown that there are high numbers of natural boosting events for serotypes 3 and 19A in children, indicating continued circulation in the vaccinated population. An observation that is most likely due to differential antibody decay to vaccine serotypes. Thus schedules which have better immunological persistence against serotype 3 and 19A after the booster dose at 1 year of age could reduce vaccine type circulation further. One such approach is currently being assessed

by PHE who are evaluating the immunogenicity of a "1+1" schedule in which a single priming dose is given in early infancy followed by a 12-month booster.[137] Such an approach has been shown to increase immunity after the 12 month booster dose for MenC vaccines when compared with the standard 2+1 schedule.[138] As this study highlights the possibility that some serotype 19A circulation could be due to antibody decay prior to boosting. Further studies which directly assess natural exposure post-priming are needed.

Serotype replacement in IPD has been realised in England and Wales with significant increases in IPD due to the high virulence serotypes 12F and 24F seen in children. Strategies to develop vaccines against these serotypes either by inclusion in increased valency conjugate vaccines or pan-pneumococcus vaccines should be a priority.

Chapter 3 Pneumococcal carriage in children from Nepal prior to PCV10 introduction assessed by conventional microbiology and DNA microarray.

3.1 Introduction

The large sample sizes needed to measure the direct protective effect of pneumococcal conjugate vaccines against invasive pneumococcal disease and/or pneumonia, especially in a resource-limited setting with poor health service infrastructure, have driven the development of alternative approaches to measure vaccine effect.[73,74,139,140] Surveillance of the pneumococcal nasopharyngeal carriage reservoir is now widely used as a proxy measure of vaccine effectiveness. Conventional, World Health Organization reviewed, pneumococcal carriage assessment is based on the isolation of single colonies of cultured pneumococcus, which are then serotyped by Quellung reaction or comparable methodologies.[76,141] This approach makes it difficult to accurately measure carriage co-colonisation with multiple pneumococcal strains because of how labour intensive the process is. This intrinsically leads to a loss of resolution in carriage detection, particularly of minor serotypes (serotypes present at lower densities), diminishes detectable vaccine effect, and results in study samples sizes needing to be larger.

3.1.1 Detecting carriage of multiple pneumococcal serotypes.

Studies which have been published since 1996 and report multiple pneumococcal serotype carriage were reviewed. Overall these studies indicate that detection of multiple serotypes is low when using conventional microbiological culture (where more than a one single colony is selected) methods (range 3.3-6.7%) compared with microarray (range 20.2-48.8%) or latex agglutination techniques (43.2%) (table 16). Comparison of different methods for detecting multiple pneumococcal serotype carriage have recently been compared in one recent study.[142]

This study found that although most methods were comparable at detecting the major serotype being carried, many of the methods (for example direct serotype detection by immunoassay) did not perform well (that is had a poor sensitivity) at detecting the additional co-colonising (minor) serotypes being carried. A DNA microarray method however, was the top performing assay for multiple serotype detection, with a reported sensitivity of 99% for detecting minor serotypes.

Table 16. ***Characteristics of studies, published since 1996, which have examined multiple pneumococcal serotype carriage.***

First author (publication year)	Region	Population	Detection method	Samples, n	Carriers, n, (%)	Carriers with >1 serotype detected, n (%)
Ussery *et al* (1996)[143]	Yukon-Kuskokwim Delta, Alaska	Children under 5 years of age	Culture	185	92 (49.7)	3 (3.3)
Lopez *et al* (1999)[144]	Oviedo, Spain	6 year old children	Culture	332	120 (36.1)	8 (6.7)
Huebner *et al* (2000)[145]	South Africa and Israel	Children aged 1-60 months	NA	NA	1899	28 (1.5)
Turner *et al* (2011)[146]	Maela, Thailand	Children aged 0-24 months	Microarray	NA	125	61 (48.8)
Turner *et al* (2011)[146]	Maela, Thailand	Children aged 0-24 months	Latex agglutination	NA	125	54 (43.2)
Turner *et al* (2011)[146]	Maela, Thailand	Children aged 0-24 months	Culture	NA	125	14 (11.2)
Kandasamy *et al* (2015)[147]	Kathmandu, Nepal	Children aged 6 weeks - 24 months	Microarray	600	297 (49.5)	60 (20.2)
Kamng'ona *et al* (2015)[148]	Blantyre and Karonga districts, Malawi	Children aged 0-13 years	Microarray	189	116 (61.4)	46 (40)

3.1.2 Evaluating pneumococcal carriage amongst Nepalese children prior to PCV introduction

In Kathmandu, Nepal there is currently a large cross-sectional study underway examining the effect that the 10-valent pneumococcal conjugate vaccine (PCV10) will have on pneumococcal carriage amongst healthy children. Between April and August of 2014, 1303 swabs were collected from children aged 6-60 months prior to introduction of PCV10 in August 2015. As part of this study the samples were processed by conventional microbiology and serotyped by Quellung reaction. Application of a DNA microarray method (whereby specific bacterial DNA sequences, such as pneumococcal capsular genes are detected) to a sub-set of these samples to detect multiple serotype carriage would thus provide a unique opportunity not only to improve the subsequent detection of vaccine effect when post-vaccine samples are collected, but also allow comparison of conventional and microarray serotyping techniques.

3.1.3 Study rationale

Therefore, this study aims to describe the pneumococcal serotype distribution amongst Nepalese children immediately prior to PCV10 introduction, and compare conventional microbiological processing and serotyping with a DNA microarray approach. This study will also identify the serotypes that are more likely to be found as a major serotype and those serotypes which are more likely to be found as minor serotypes.

3.2 Methods

3.2.1 Study design and participants

As part of the parent study, described above, 1303 children were recruited between April 2014 and August 2014. The first 437 of these children, recruited between April and May 2014, were included in the study assessing microarray detection of pneumococci. Parents/guardians of healthy Nepalese children aged 6-60 months attending the paediatric outpatient department at the Patan Hospital, Kathmandu were approached for enrolment of their child. Children who were afebrile and in good health, as determined by medical history, were included in the study. Children who were febrile or had a recent respiratory illness were temporarily excluded until a later date. In this pre-PCV cohort children who had received a dose of PCV were excluded. After informed consent by the child's parent and/or guardian, demographic data and a nasopharyngeal swab were collected. The primary objective of this study was to determine the serotype distribution of pneumococci carried amongst Nepalese children prior to PCV10 introduction using a DNA microarray. Secondary objectives included; comparison of conventional serotyping with DNA microarray serotyping, determining the proportion of children carrying a PCV10 serotype, and identifying which pneumococci are predominantly detected as major and minor serotypes. Ethical approval was granted from the Oxford Tropical Research Ethics Committee (OxTREC, 28-14) and the Nepal Health and Research Committee (NHRC, 39/2014).

3.2.2 Specimen collection and conventional microbiological processing

Nasopharyngeal swabs were collected and processed by conventional microbiological methods according to WHO guidelines.[76] That is, a single nylon flocked nasopharyngeal swab (Thermo Fisher Scientific, Leicestershire, United Kingdom) was collected from each participant, and placed in skim-milk-tryptone-glucose-glycerin (STGG) medium. Swabs were then kept at 2-8

ºC until they were processed, within eight hours, in the laboratory. Swabs were streaked onto 5% sheep blood agar and the plates incubated at 37 ºC overnight. Following streaking out onto agar all swabs were then stored at -70 ºC. Mucoid, alpha-haemolytic colonies were selected from the agar plate and sub-cultured on 5% sheep blood agar with an optochin disc. If more than one morphologically distinct colony was present, then these were sub-cultured out individually. Optochin sensitive strains were then stored at -70 ºC in Microbank bacterial storage tubes (Thermo Fisher Scientific, Leicestershire, United Kingdom) until serotyping was conducted.

3.2.3 Quellung serotyping of pneumococci

After overnight culture on 5% sheep blood agar, pneumococcal strains were serotyped using the Quellung test and phase contrast microscopy.[15] On a microscope slide pneumococcal colonies were emulsified in 1 µl of saline before 1 µl of pneumococcal anti-serum (Statens Serum Institut, Hillerod, Denmark) was added to the mixture. The wet preparation was viewed under 100X magnification using oil immersion. Preparations were classified as positive (capsular swelling and bacterial aggregation) or negative reactions by trained technicians. The final serotype was determined by working through the serotyping pathway described by the Statens Serum Institut.[15] Pneumococci were classified at non-typeable if there was a negative reaction to the Omni pneumococcal antisera.

3.2.4 Culture of STGG dilutions, DNA extraction, and microarray analysis

Those specimens that had pneumococcus identified by conventional microbiology were selected for microarray processing using the BUGS Senti-SP microarray. Initially STGG from each sample was plated onto Streptococcal selective COBA plates (Oxoid, Hampshire, United

Kingdom) in neat, 1:10, and 1:100 dilutions. After overnight incubation at 37 ºC, the plate with the highest density of non-confluent growth was then selected, colonies from across the entire plate were scraped into 1 ml of saline and stored at -20 ºC or below until processed for DNA extraction.

Genomic DNA were extracted from pneumococcal plate sweep suspensions by BUGS bioscience, using QIAamp DNA Mini Kit protocol (Qiagen, Germany). The suspensions were centrifuged to pellet the bacterial, re-suspended in 180 µL freshly prepared lysis buffer (20mg/mL lysozyme, 1mM Tris-HCL, 500 mM EDTA, 10% Triton) and incubated at 37ºC for 60 min. Proteinase K (20 µL) and Qiagen buffer AL (200 µL) were added to the mixture and incubated for another 60 min at 56 ºC. 4 µL RNase A was added and incubated at room temperature for 5 min before incubation at 70 ºC for 10 min. The standard manufacturer's extraction protocol was followed after lysis and the extracted DNA stored at 4ºC or below for microarray analysis.

Molecular serotyping was conducted using the Senti-SP microarray (BUGS Bioscience, London, United Kingdom; http://bugsbio.org).[149] DNA samples were fluorescently labelled and hybridized to the Agilent 8×15K format microarray according to manufacturer's instructions for the Agilent genomic DNA ULS labelling and oligo aCGH hybridisation reagent kits. Microarray data were statistically analysed using a Bayesian hierarchical model to determine the serotype, or combination of serotypes, present in the sample and assign a relative abundance to each serotype detected. Additional components of the microarray enable detection of non-pneumococcal *Streptococcus spp.,* antibiotic resistance genes, and assess genetic relatedness.

3.2.5 Statistical analysis

The ratio of major to minor serotypes was calculated by dividing the number of times a serotype was detected by DNA microarray as the most abundant serotype within a sample by the number of times it was detected as a lesser abundant serotype within a sample.

Proportions were compared using Fisher's exact test. The Welch Two Sample t-test was used for comparing means. All statistical analyses were performed using R version 3.3.1.[134]

3.3 Results

Between April and May 2014, 437 children were enrolled into the parent study. Of these 316/437 (72.3%) were found to be carrying pneumococcus, following processing of nasopharyngeal swabs by conventional microbiology. There were no differences in basic demographic characteristics collected at the time of enrolment when comparing children who were pneumococcal carriers with those who were non-carriers (table 17).

Table 17. ***Characteristics of participants enrolled in the study.***

Characteristic	All, N=437	Carriers, N=316	Non-carriers, N=121	p (Carriers versus non-carriers)
Mean age, years (SD)	1.7 (1)	1.7 (1)	1.8 (1)	0.2343
Males, N (%)	261 (59.7)	188 (59.5)	73 (60.3)	0.9134
Term delivery, N (%)	393 (89.9)	280 (88.6)	113 (93.4)	0.1576
Mean birth weight, kg (SD)	3 (0.6)	3 (0.6)	3 (0.5)	0.1692
Number of people living at home, N (SD)	4.7 (2.2)	4.7 (2.3)	4.6 (2)	0.5288

3.3.1 Serotype-specific distribution of pneumococci carried amongst Nepalese children as determined by conventional microbiology

Of the 316 pneumococcal carriers identified by conventional culture, four (4/316, 1.3%) had multiple serotype carriage, with two distinct pneumococcal serotypes identified from each of these four samples. Non-typeable (42/316, 13.3%) followed by serotype 19F (33/316, 10.4%) pneumococci were the most frequently identified serotypes amongst pneumococcus positive swabs (figure 5). 108/316 (34.2%) of pneumococcal carriers had a PCV10 serotype identified.

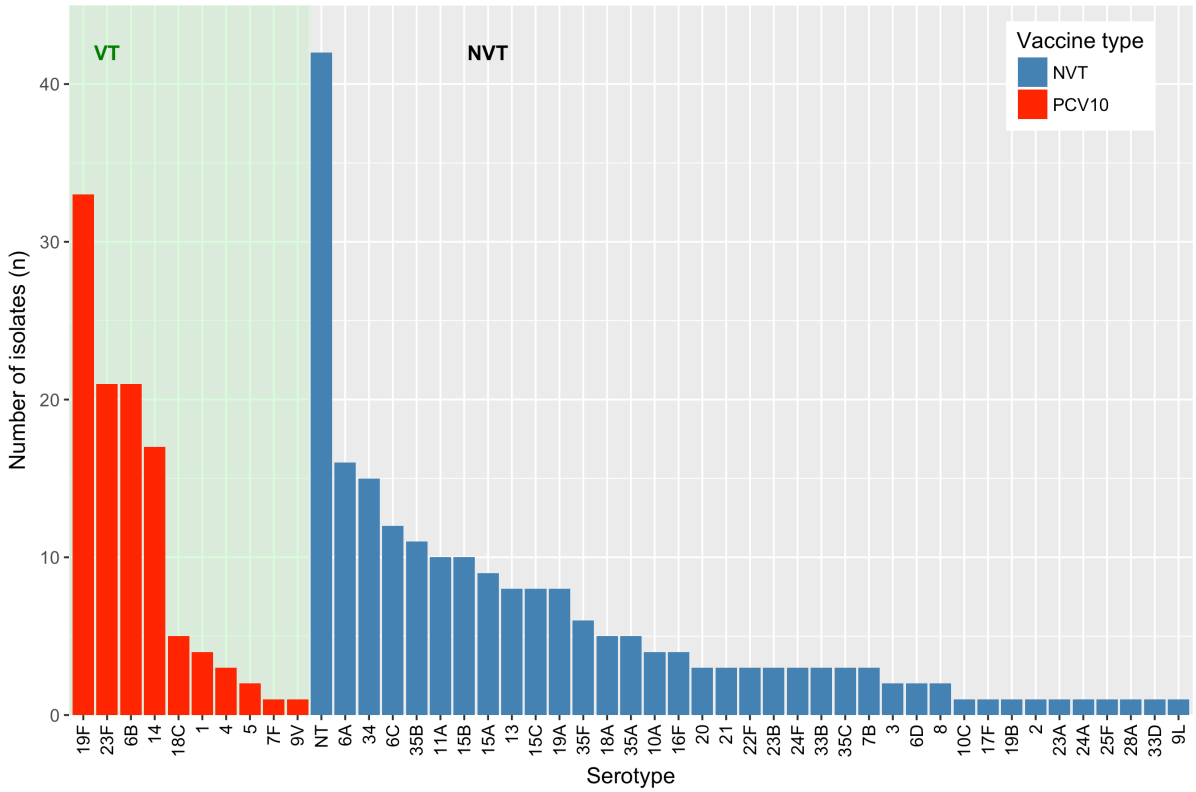


Figure 5. ***Serotype-specific distribution of pneumococci identified by conventional microbiology and Quellung reaction amongst nasopharyngeal swabs collected from healthy Nepalese children prior to PCV10 introduction.***
Swabs were collected between April and May 2014 from children attending the outpatient's department of Patan Hospital, Kathmandu, Nepal. PCV10 serotypes are coloured red and non-vaccine types (NVTs) are coloured blue.

3.3.2 Serotype-specific distribution of pneumococci carried amongst Nepalese children as detected by DNA microarray

Serotyping by DNA microarray was conducted on the 316 participant samples which had pneumococcus identified by conventional microbiological techniques. Pneumococcus was not detected amongst 6/316 samples with five of these samples having a closely related non-pneumococcal *Streptococcus spp.* detected instead. One sample had no detectable *Streptococcus spp.* by the microarray. A total of 509 distinct pneumococcal serotypes and 93 closely related non-pneumococcal *Streptococcus spp.* were detected amongst the 316 swabs analysed. 138/316 (43.7%) swabs were found to have more than one pneumococcal serotype detected. Non-typeable followed by serotype 19F pneumococci were the most frequently

identified serotypes, being found on 74/509 and 39/509 occasions, and amongst 68/316 and 39/316 analysed swabs respectively (figure 6).

Across all of the pneumococci detected, PCV10 serotypes were identified on 161/509 (31.6%) occasions. PCV10 serotypes were detected amongst 137/316 (43.4%) of the swab samples analysed. Of these 137 isolates the PCV10 serotypes were detected as a minor serotype on 34/137 (24.8%) of those swabs.

The tendency of a pneumococcal serotype to be detected as a major as opposed to minor serotype can be assessed by comparing the frequency that each serotype is found as either a major or minor serotype (figure 7). Ten serotypes (serotypes 2, 6D, 8, 9L, 9V, 18F, 20, 22A, 28A, and 39) were detected as major serotypes but not as minor serotypes. Serotype 15A had the highest determinable ratio of major to minor detections, being found eight times as a major serotype and only once as a minor (table 18). Five serotypes (serotypes 7A, 7C, 27, 28F, and 29) were only found as minor and not as major serotypes and thus had a major to minor serotype ratio of zero. Serotype 23A had the lowest non-zero major to minor serotype ratio of 0.25, being found once as a major serotype and on four occasions as a minor serotype.

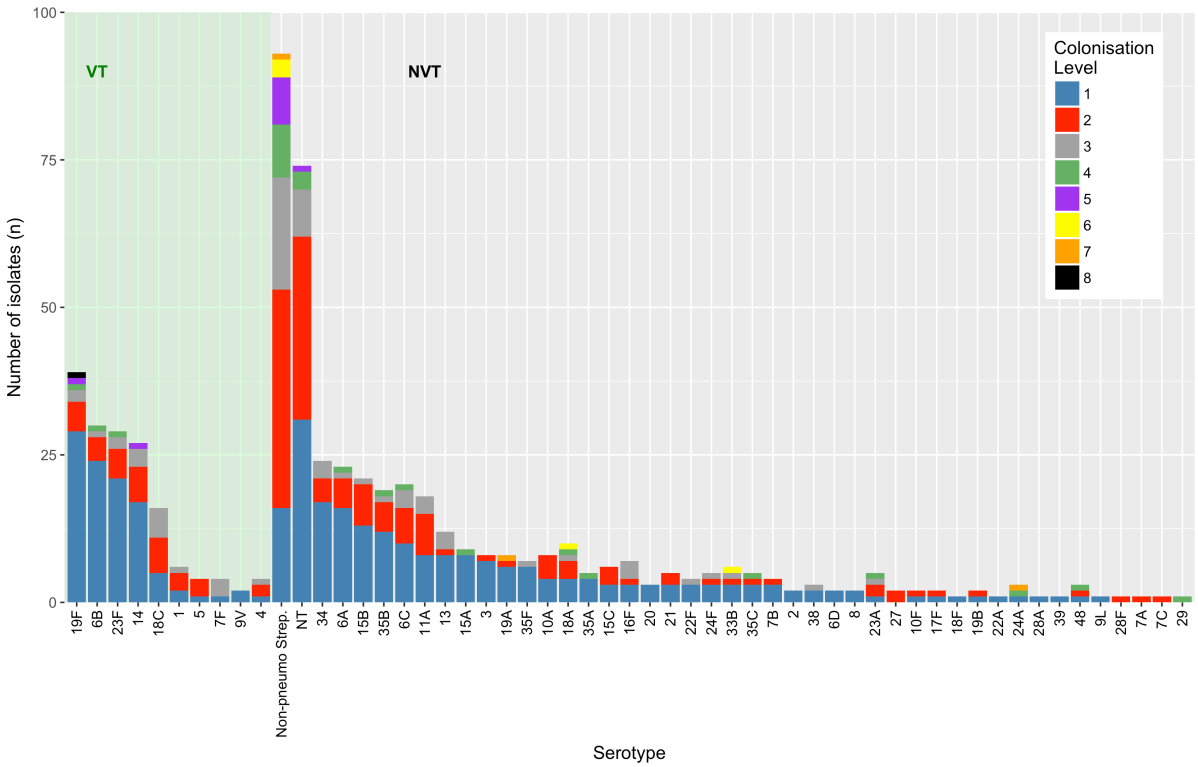


Figure 6. ***Serotype-specific distribution of pneumococci identified by DNA microarray amongst nasopharyngeal swabs collected from healthy Nepalese children prior to PCV10 introduction.***
Swabs were collected between April and May 2014 from children attending the outpatient's department of Patan Hospital, Kathmandu, Nepal. 316 swabs which had pneumococcus identified by conventional microbiology were analysed using DNA microarray. In addition to detecting the serotype of each pneumococcus, the microarray also detects closely related *Streptococcus spp.* which are represented in this graph as 'Non-pneumo Strep.' Analysis of the microarray signature also allows a relative abundance of each serotype to be calculated. Those serotypes which were found to be the most abundant on a swab were classified at having a colonisation level of 1. The subsequent colonisation levels are then assigned to the next most abundant serotypes detected from the child's swab.

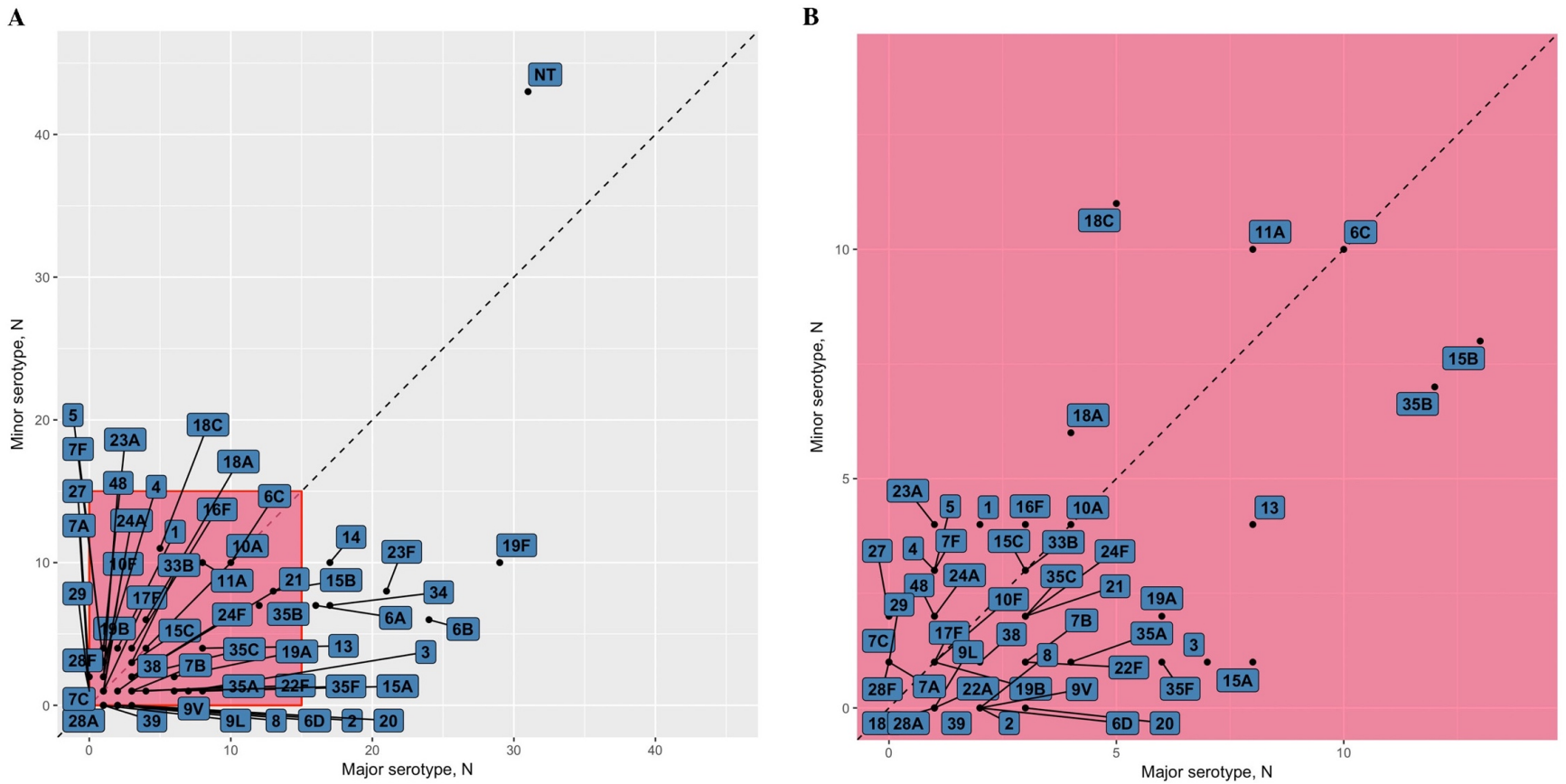


Figure 7. ***Frequency of isolation as a major compared with a minor serotype for each of the pneumococcal serotypes detected amongst Nepalese children prior to PCV10 introduction.***

Pneumococcal serotypes were detected using a DNA microarray which also reports the relative abundance of each detected serotype from a single sample. The most abundant serotype is classified as the major serotype and any subsequent serotypes classified as minor serotypes. All serotypes are displayed in panel A, whilst those serotypes which were only detected at lower frequencies (red square) are focused upon in panel B.

Table 18. ***Major to minor serotype ratio of pneumococci detected by microarray.***

Serotype	Major, n (N=319)	Minor, n (N=283)	Ratio (95% CI)	p
NT	31	43	0.7 (0.4-1)	0.0467
19F	29	10	2.9 (1.3-6.4)	**0.0073**
6B	24	6	4 (1.5-11.4)	**0.0023**
23F	21	8	2.6 (1-6.4)	**0.0360**
34	17	7	2.4 (0.9-6.4)	0.0945
14	17	10	1.7 (0.7-3.8)	0.3280
6A	16	7	2.3(0.8-6.1)	0.1357
Non-pneumo. Strep.	16	77	0.2 (0.1-0.3)	**<0.0001**
15B	13	8	1.6 (0.6-4.1)	0.5063
35B	12	7	1.7 (0.5-4.7)	0.4850
6C	10	10	1 (0.3-2.4)	0.8229
15A	8	1	8 (1-322.5)	**0.0407**
13	8	4	2 (0.5-8.2)	0.3933
11A	8	10	0.8 (0.2-2)	0.4825
3	7	1	7 (0.8-285.8)	0.0724
35F	6	1	6 (0.6-249.2)	0.1275
19A	6	2	3 (0.5-27.5)	0.2929
18C	5	11	0.5 (0.1-1.2)	0.1253
35A	4	1	4 (0.4-176.8)	0.3774
10A	4	4	1 (0.2-4.8)	1
18A	4	6	0.7 (0.1-2.5)	0.5277
20	3	0	-	0.2514
22F	3	1	3 (0.2-140.9)	0.6264
7B	3	1	3 (0.2-140.9)	0.6264
21	3	2	1.5 (0.2-16.1)	1
24F	3	2	1.5 (0.2-16.1)	1
35C	3	2	1.5 (0.2-16.1)	1
15C	3	3	1 (0.1-6.7)	1
33B	3	3	1 (0.1-6.7)	1
16F	3	4	0.8 (0.1-4)	0.7117
2	2	0	-	0.5010
6D	2	0	-	0.5010
8	2	0	-	0.5010
9V	2	0	-	0.5010
38	2	1	2 (0.1-105.3)	1
1	2	4	0.5 (0-3.1)	0.4273
18F	1	0	-	1
22A	1	0	-	1
28A	1	0	-	1
39	1	0	-	1
9L	1	0	-	1
10F	1	1	1 (0-69.8)	1
17F	1	1	1 (0-69.8)	1
19B	1	1	1 (0-69.8)	1
24A	1	2	0.5 (0-8.5)	0.6033
48	1	2	0.5 (0-8.5)	0.6033
4	1	3	0.3 (0-3.7)	0.3465
5	1	3	0.3 (0-3.7)	0.3465
7F	1	3	0.3 (0-3.7)	0.3465
23A	1	4	0.25 (0-2.2)	0.1925
27	0	2	0 (0-4.7)	0.2206
28F	0	1	0 (0-34.6)	0.4701
7A	0	1	0 (0-34.6)	0.4701
7C	0	1	0 (0-34.6)	0.4701
29	0	1	0 (0-34.6)	0.4701

*Non-pneumo. Strep. = non-pneumococcal *Streptococcus spp.* detected by microarray.

3.3.3 Comparison of conventional culture with microarray for the detection of pneumococci

A total of 320 (312 single serotypes from 312 swabs and four swabs with two serotypes detected on each) distinct isolates were detected by Quellung. On 240/320 (75%) occasions the serotype identified by conventional microbiology and Quellung was the same as the major serotype identified by DNA microarray. The isolates detected by microarray included the same serotype (either major or minor) as conventional microbiology with Quellung serotyping on 277/320 (86.6%) occasions (figure 8). 42 samples had a Quellung determined serotype which did not match a serotype detected by microarray processing of the same sample; 13/42 (31%) of these were matches within the same sero*group*, 18/42 (42.9%) were classified as NT by Quellung (10/18, 55.6%, of these had a non-pneumococcal *Streptococcus spp.* detected by microarray), and 11/42 (26.2%) samples were completely discrepant in the serotypes detected.

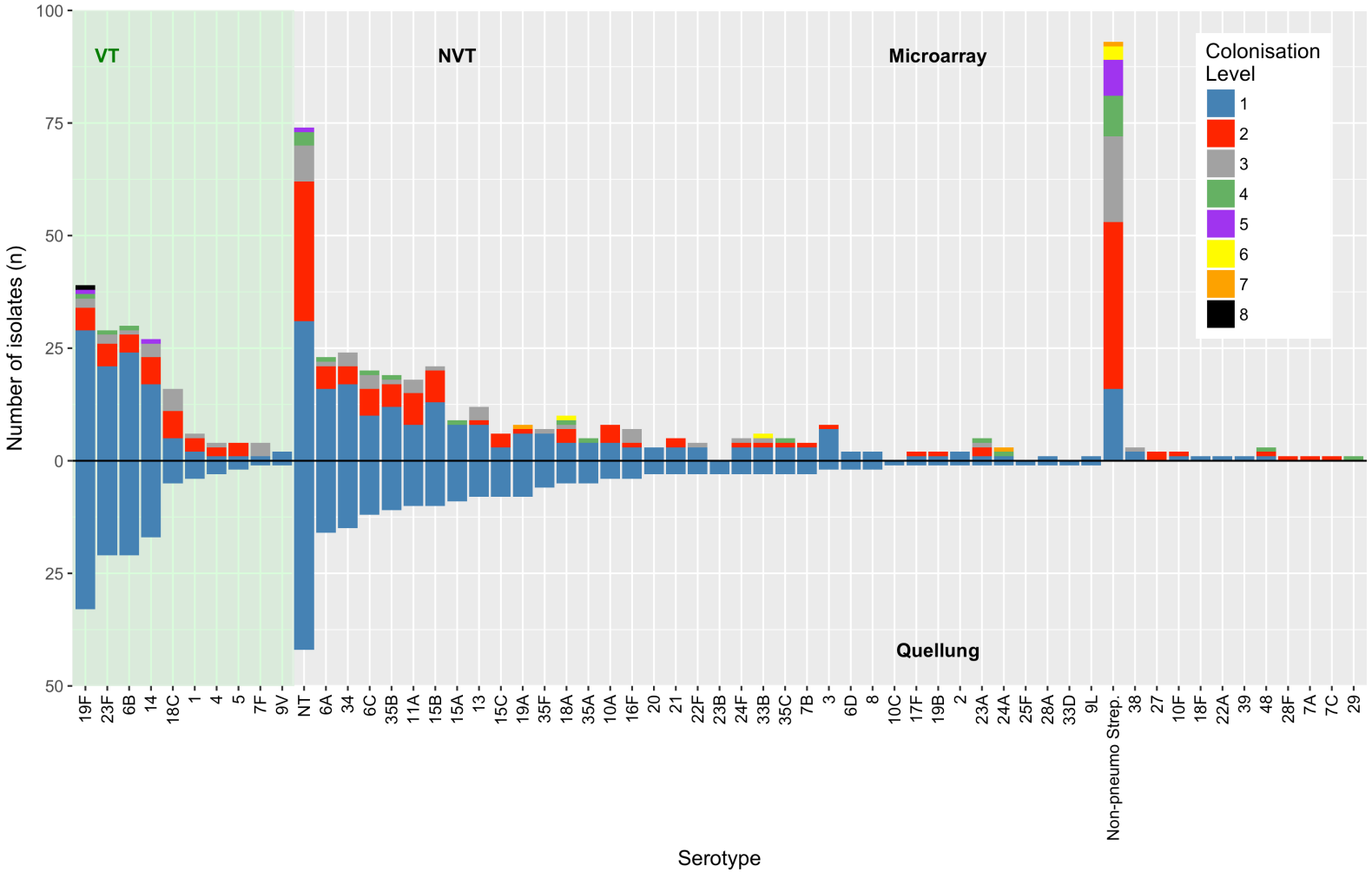


Figure 8. ***Serotype-specific distribution of pneumococci identified by DNA microarray (bars on the top) compared with conventional microbiology and Quellung (bars on the bottom), and ordered by Quellung serotype prevalence, amongst nasopharyngeal swabs collected from healthy Nepalese children prior to PCV10 introduction.***
Serotype detected by microarray was concordant with serotype detected by conventional culture with Quellung serotyping on 277/320 (86.6%) occasions

3.4 Discussion

This is the largest study to date to make a comparison of conventional microbiology with DNA microarray for the detection of multiple serotype carriage in children. It shows that there is good concordance between conventional microbiology and microarray techniques for the detection of the same serotypes from each sample. This study also shows that the application of this microarray technique improves the detection of minor serotypes, which in turn also increases the overall detection of vaccine serotypes from children's swabs, and so would improve the assessment of PCV efficacy across populations. Application of this DNA microarray to the assessment of vaccine efficacy studies should be considered instead of conventional microbiology, particularly where it is important to minimise sample size.

3.4.1 Serotype distribution of pneumococci amongst Nepalese children prior to PCV10 introduction.

Using the microarray method, the most frequently detected (by microarray) PCV10 covered sertoypes in this present study (in descending order of prevalence) were serotypes 19F, 6B, 23F and 14. This is consistent with a prior study conducted in the same region four years' prior which showed that serotype 14, followed by serotypes 19F, 23F, and 6B were the most frequently detected PCV10 serotypes.[150] Notably however, there has been a change in the order of which of these serotypes were detected most frequently. This may be due to natural shifts in prevalence due to competition between pneumococcal strains, or it may be due to chance associated with the cross-sectional population sample enrolled for each of the two studies.

With regard to serotypes included in PCV13 but not PCV10, serotype 3 was infrequently detected, whilst serotypes 6A and 19A were noted to be present at moderate levels. Given that PCV10 is being utilised in this setting now, it is particularly important to follow whether there is any significant serotype replacement due to these additional serotypes, which have been

associated with significant amounts of IPD in other settings, and could be covered by the use of PCV13.[90]

3.4.2 Conventional culture and Quellung versus DNA microarray for detecting and serotyping pneumococci.

The major serotype detected by microarray was the same serotype detected by conventional microbiology and Quellung amongst 75% of the samples. The observed difference between the two techniques may be because some pneumococcal serotypes, although present at a lesser abundance on the culture plate, are easier to identify than other serotypes. For example, the large mucoid colonies typically formed by serotype 8 pneumococci compared with the small dry colonies formed by NT pneumococci.

The agreement between the two techniques for identifying at least one common serotype was 86.6% and is in keeping with a prior study which reported an agreement of 87.2% between the same microarray technique and conventional microbiology with Quellung.[151] Relaxing the level of agreement to the sero*group* level gives a concordance of 90.6% between the two techniques for detecting the same pneumococcal sero*group*. Again this is keeping with a prior study comparing the same techniques which reported an agreement of 95.2% at the sero*group* level.[151]

Significantly more pneumococci were detected from samples using microarray than in those processed conventionally (509/316 vs 320/316, $p<0.0001$), with multiple-serotype carriage being detected frequently (43.7%) amongst swabs analysed by microarray. This is in keeping with a prior study which reported 48.8% of samples having multiple-serotypes detected.[151] However, this level of multiple-serotype detection is higher than a prior study in Nepalese children using microarray detection which reported multiple-serotypes to be present amongst 20.2% of pneumococcus positive swabs.[150] This apparent difference may be due to the fact that the children in this present study had an older average age than those in the prior study, and so

age related differences in multiple-serotype carriage may be a factor. In addition, the difference in multiple-serotype carriage may be due to advances in microarray technology with the most recent platform having the ability to detect up to eight distinct serotypes compared with only five for the earlier version used in the prior study.

Significantly more isolates with a PCV10 serotype were detected amongst children's swabs by microarray compared with conventional processing (108/316 versus 137/316, p=0.0222). As such, if the presence of a PCV10 serotype on a child's swab were to be used as an endpoint for a vaccine efficacy study, the microarray assay would provide a more powerful approach for detecting vaccine effect.

3.4.3 Strengths and weaknesses

The strengths of this study are: the size of the cohort across which the comparison of conventional culture and Quellung serotyping with microarray detection has been conducted; the potential for further increasing the number of samples analysed by assessing the remaining participants in the parent cohort; the ability to meta-analyse these data with those from the prior study conducted in the same setting which also used a comparable technique; and the detailed baseline which this study provides for measuring vaccine efficacy should this data set be compared with a post-vaccine cohort.

Weaknesses of this study include the short time-period over which samples were collected, and the fact that processing of the conventionally analysed samples was not done in parallel with those that were analysed by microarray. Seasonal variations in overall pneumococcal carriage prevalence amongst a population occur, whilst natural shifts in serotype distribution due to competition between pneumococcal strains are also at play at any given time. As such the short time-period over which the samples included in this analysis were collected introduces a higher chance of these data being influenced by such factors, as opposed to a cohort where samples are collected over a longer time-period across which such confounding factors would more

likely be averaged out. The fact that conventional processing was not done in parallel with microarray processing introduces the possibility that there were variations introduced due to the re-culturing or the storage process.

3.4.4 Conclusions

This study demonstrates that pneumococcal carriage is prevalent amongst Nepalese children in the Kathmandu Valley and that nearly half of those children who carry pneumococcus, carry more than one serotype. PCV10 serotypes are frequently carried in this population, whilst detection of these serotypes is improved by microarray processing. Assessment of the pneumococcal serotype distribution in this population following the introduction of PCV10, ideally using a microarray approach to serotype detection, will provide a powerful insight into vaccine efficacy.

Chapter 4 Developing a non-invasive molecular diagnostic and clinical decision-making tool for pneumococcal pneumonia.

4.1 Introduction

Given the large global burden of pneumococcal disease, the development of an accurate diagnostic tool for pneumococcal pneumonia should take a high priority in the scientific and medical community. Improved diagnostics would allow quicker and more directed treatment, decrease length of hospital admission, and would improve estimations of disease burden.

4.1.1 Limitations of current diagnostic tests for pneumococcal pneumonia

Conventional microbiology could have a role in those cases of pneumonia which are associated with a pleural effusion and/or empyema should there be a clinical need to sample the para-pneumonic fluid. However, not all cases of pneumonia are associated with an effusion or empyema, with maximal estimates describing 57% of pneumonia cases having fluid in the para-pneumonic space.[37] And, although conventional microbiological testing of pleural fluid is better when compared with blood cultures (because not all cases of pneumonia result in bacteraemia) for the detection of a causative organism, it is still less than ideal. For example, one study showed that a causative organism was identified in 29% of pleural effusion and 62% of empyema cases, whilst another study reported a causative organism being detected in 58.9% of pneumonia cases with complicated para-pneumonic effusion.[45,152] Application of molecular based techniques to the assessment of para-pneumonic samples can improve the diagnostic yield.[48,49] However, this approach is still undermined by the fact that sampling of pleural fluid is not standard practice for the diagnosis and management of pneumonia, is invasive, and can result in life-threatening complications.

Urine antigen testing for the pneumococcal C-polysaccharide has been shown to be a useful tool in adults for the diagnosis of pneumococcal disease.[53] However, in children the much higher prevalence of asymptomatic carriage compared with adults, means the assay is unable

to differentiate healthy children with asymptomatic carriage from those with disease.[54,55] Attempts have been made to detect pneumococcal capsular polysaccharide in urine by ELISA however it was found that the levels of capsular polysaccharide in the urine were extremely variable, fluctuating according to disease severity and pneumococcal serotype.[153]

Using serum antibody responses to pneumococcal antigens, such as pneumococcal surface adhesin A (psaA) and capsular polysaccharides, for diagnosing pneumococcal pneumonia have previously been assessed. Studies showed that the levels of antibody to psaA had good sensitivity (85-89%) and specificity (83-98%) in adults, however another study showed that this unfortunately was not the case in children (sensitivity 42% and specificity 97%).[58,154,155] Antibody responses to capsular polysaccharides have also been used epidemiologically to assess the cause of pneumonia in children, with studies showing a good correlation of antibody responses to pneumococcal capsule with detection of pneumococcus by blood or pleural fluid culture.[156,157] However, asymptomatic carriage of pneumococcus also generates antibody responses to capsular polysaccharides, and thus the utility of this assay as a diagnostic test is poor.

4.1.2 Using carriage density to diagnose pneumococcal pneumonia.

Real time PCR of the autolysin gene (*lytA*) has been applied to clinical specimens (such as nasopharyngeal swabs or sputum) in order to quantitate the density of pneumococci present from the site of sampling. These data can then be related to clinical measures such as composite diagnostic tests for pneumococcal pneumonia. For example, one study used a combination of chest x-ray, blood culture, sputum culture, and urine antigen testing to define participants with likely pneumococcal pneumonia.[59] Review of the literature demonstrated four cohorts of adults which had different clinical specimens assessed for pneumococcal density using real-time PCR for *lytA* (table 19). Each of these studies determined the sensitivity and specificity of using a

threshold value of pneumococcal density for diagnosing pneumococcal pneumonia through the generation of receiver operator curves. One cohort which used nasopharyngeal swabs and another which used sputum samples reported sensitivities and specificities that would be high enough for consideration as a clinical test. One recent study assessing carriage density in children admitted with pneumonia was found.[158] This study defined pneumococcal pneumonia as those cases where pneumococcus was microbiologically confirmed by culture of blood or pleural fluid. This study showed a carriage density of nearly 8 million *lytA* copies/ml had limited sensitivity (64%), but good specificity (92%), for distinguishing microbiologically confirmed pneumococcal pneumonia from controls.

Table 19. ***Studies which have examined pneumococcal density using real time lytA PCR as a diagnostic tool for pneumococcal pneumonia.***

Publication first author (year)	Region	Population	Participants with pneumococcal pneumonia, n	Specimen	Threshold	Sensitivity	Specificity
Albrich *et al* (2012)[59]	South Africa	HIV infected adults admitted to hospital with pneumonia	76/280	Nasopharyngeal swab	≥8000 copies/ml	82.2%	92%
Albrich *et al* (2014)[159]	South Africa	HIV infected adults admitted to hospital with pneumonia	81/222	Induced sputum	≥10000 copies/ml	78.1%	80%
Stralin *et al* (2014)[160]	Sweden	Adults admitted with radiologically confirmed pneumonia	32/235	Sputum	≥100000 copies/ml	94%	96%
Stralin *et al* (2014)[160]	Sweden	Adults admitted with radiologically confirmed pneumonia	32/235	Nasopharyngeal aspirate	≥100000 copies/ml	53%	93%
Baggett et al (2017)[158]	Bangladesh, The Gambia, Kenya, Mali, South Africa, Thailand, and Zambia	Children under 5 years of age admitted pneumonia	56/4035	Nasopharyngeal swab	>7943282 copies/ml	64%	92%

4.1.3 Surveillance of children admitted with pneumonia at Patan Hospital, Kathmandu, Nepal

As part of a study which is assessing the effect of PCV10 on admissions for paediatric pneumonia at Patan Hospital, Kathmandu, Nepal, nasopharyngeal swabs are collected and processed by conventional microbiological techniques.[76] In addition, detailed clinical data which includes, changes present on chest x-rays, white blood cell and neutrophil counts, length of illness, and length of admission are collected. These samples and data therefore provide a unique opportunity to assess whether pneumococcal carriage density in children can be used to facilitate diagnosis of pneumococcal pneumonia.

4.1.4 Rationale for a study

Therefore, this study aims to use the nasopharyngeal samples collected from children admitted with pneumonia at Patan Hospital to quantitate the density of pneumococci using real time PCR. These data can then be used to determine if density relates to pneumococcal pneumonia (as classified by a composite diagnostic standard) and for comparing PCR with conventional culture for detection of pneumococcus.

4.2 Methods

4.2.1 Study design and definition of pneumococcal pneumonia

Between March 2014 and August 2016, Nepalese children admitted to Patan Hospital with clinician diagnosed pneumonia were enrolled into the parent study from which samples were analysed in this present study. A random selection of 368 participants were included in this current study.

Based on surveillance data over the last ten years, the three most frequently isolated serotypes in the context of paediatric invasive pneumococcal disease at Patan Hospital are serotypes 1, 5, and 14 (see chapter 1 figure 1). As such children with an abnormal chest x-ray, a blood neutrophil level over the upper normal limit (over $7.5x10^9$), and carrying either serotypes 1, 5, or 14 in their nasopharynx were classified as having pneumococcal pneumonia. In addition, any child with an abnormal chest x-ray and blood culture positive for pneumococcus was classified as having pneumococcal pneumonia.

The primary objective of this study was to determine the diagnostic accuracy of qPCR for *lytA* from nasopharyngeal swabs for differentiating pneumococcal pneumonia from non-pneumococcal pneumonia in Nepalese children. Secondary objectives included; comparing standard microbiological detection of pneumococcus with detection by *lytA* qPCR, and comparing pneumococcal carriage densities between children with pneumococcal pneumonia and those with non-pneumococcal pneumonia. Ethical approval was granted from the Oxford Tropical Research Ethics Committee (OxTREC, 09-15) and the Nepal Health and Research Committee (NHRC, 286/2014).

4.2.2 Study participants

Children aged 2 months to 14 years who were admitted to Patan Hospital with clinician diagnosed pneumonia were approached by trained research staff for enrollment in the parent

study. Children who had any significant severe disease or disorder that may influence the result of the study, were excluded from participation. Following completion of informed consent by the child parent/guardian demographic and clinical data (including chest x-ray) were collected. In addition, clinical samples including; nasopharyngeal swab, blood, and urine were collected.

4.2.3 DNA extraction from STGG media.

Nasopharyngeal swabs in STGG media which had been stored at -80 °C were defrosted. DNA was extracted from 200 µl STGG media within which the nasopharyngeal swab had been stored using a modified protocol based around the QIAGEN DNeasy 96 kit (QIAGEN, Manchester, United Kingdom).

The STGG was mixed in microtubes with 100 µl of pre-lysis solution consisting of; 20 mg/ml of chicken egg lysozyme (Fisher Scientific, Leicester, United Kingdom), 50 Units/ml of mutanolysin (Sigma-Aldrich, Dorset, United Kingdom), and Tris EDTA buffer (Sigma-Aldrich, Dorset, United Kingdom) to make up the remaining volume. Samples were then incubated for one hour at 37°C and then were centrifuged for 30 seconds at 3800 g to remove any condensation from the caps. 20 µl of protein kinase and 50 µl of AL buffer was then added to each sample before mixing and then centrifuging for 30 seconds at 3800 g. Samples were then incubated for one hour at 56 °C and then centrifuged for 30 seconds at 3800g. 150 µl of AL buffer was then added to each sample before incubating at room temperature for 10 minutes. 250 µl of absolute ethanol was then added to each sample before mixing and then transferring each sample to a well on a DNeasy 96 well plate. Plates were then centrifuged for ten minutes at 3800 g. 500 µl of AW1 buffer was then added to each well and the plates centrifuged for five minutes at 3800 g. 500 µl of AW2 buffer was then added to each well and the plates centrifuged for fifteen minutes at 3800 g. 100 µl of AE buffer warmed to 37 °C was then added to each well, plates incubated at room temperature for five minutes and then

centrifuged for 2 minutes at 3800 g and the eluted DNA collected. A further 100 µl of AE buffer was again added to each sample well and the plates centrifuged for 2 minutes at 3800 g and the eluted DNA collected. Eluted DNA was stored at -20 °C or below.

4.2.4 *lytA* real time PCR

Assays were carried out in a final reaction volume of 25 µl and performed in duplicate. Each assay consisted of; previously described pneumococcal autolysin primers and probe (Life Technologies Limited, Paisley, United Kingdom) (table 20), 12.5 µl of Taqman gene expression mastermix (Thermo Fisher Scientific, Leicestershire, United Kingdom), 10 µl of template DNA, and 1µl of molecular grade water (Thermo Fisher Scientific, Leicestershire, United Kingdom).[161]

Table 20. ***Pneumococcal autolysin (lytA) primers used to assess pneumococcal density***

Forward primer	5'-acgcaatctagcagatgaagca-3'
Reverse primer	5'-tcgtgcgttttaattccagct-3'
Probe	5'-tgccgaaaacgcttgatacagggag-3' (5' FAM; 3' MGB)

A plasmid containing *lytA* (GenExpress, Berlin, Germany) was used in defined concentrations (ranging from 10^2-10^8 copies per/ml) to generate a standard curve from which the total copy number of *lytA* could be determined for each sample.

All samples were run in duplicate on a StepOnePlus real time PCR platform (Thermo Fisher Scientific, Leicestershire, United Kingdom). Samples underwent a single first stage at 95 °C for 10 minutes followed by second stage consisting of 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute.

4.2.5 Statistical analysis.

Initial analysis of qPCR results was performed using the StepOnePlus Software V2.3.[162] Using this software the baseline was set to automatic and the threshold set at 0.05. Criteria for a plate passing were; the standard curve had a slope between -3.1 and -3.6, the standard curve had a correlation greater than 0.9, negative controls were negative, and positive controls showed amplification. For each participant the mean copy number of the duplicate readings were calculated.

Proportions were compared using Fisher's exact test. Means were compared using the Welch Two Sample t-test. Mean gene copies were log10 transformed before comparison using the Welch Two Sample t-test. If there were no detectable gene copies for a participant, then they were assigned an arbitrary value of 0.5 before log10 transformation.

ROC curves, area under the curve (AUC), sensitivities and specificities were derived using the pROC package in R.[163] Threshold values were chosen by identifying the number of *lytA* gene copies/ml which related to the sensitivity and specificity values that had the least difference. Sensitivity analyses were conducted; to examine the effect of removing serotype 14 carriage from the diagnostic criteria, and to examine the difference between children with pneumococcal pneumonia carrying serotypes 1, 5, or 14 pneumococci and children with normal chest x-rays also carrying serotypes 1, 5, or 14 pneumococci (appendix C figure 42 and appendix C figure 43).

All analyses that were not performed using the StepOnePlus Software were conducted using R version 3.3.1.[134]

4.3 Results

Between March 2014 and August 2016, 793 children admitted with pneumonia to Patan Hospital, Kathmandu, Nepal were recruited into the parent study. Of these 368 randomly chosen children with clinically diagnosed pneumonia had real-time PCR for pneumococcal *lytA* performed. 179/368 (48.6%) children had an abnormal chest x-ray, as defined by either the detection of end-point consolidation or infiltrates (table 21). Children with radiologically confirmed changes on their chest x-rays were significantly older than those who had no abnormalities detected radiologically. Length of illness, length of admission, and blood cell markers were significantly higher in children with radiological changes compared with those that did not have any radiological changes.

Table 21. ***Characteristics of patients from whom nasopharyngeal swab samples were collected.***

Characteristic	All admitted children (N=368)	Children with radiologically confirmed pneumonia (N=180)	Children without radiologically confirmed pneumonia (N=181)	p (radiologically confirmed versus non-radiologically confirmed)
Age, median years (IQR)	1.4 (0.6-2.7)	1.8 (0.8-3.8)	1.1 (0.5-2.2)	**<0.0001**
Males, n (%)	218 (59.2%)	98 (54.4%)	115 (63.5%)	0.0873
Weight, median kg (IQR)	9 (7.5-12)	10 (8-12)	8.9 (7.4-11)	**0.005**
Had antibiotics prior to admission, n (%)	180 (48.9%) (n=2, missing data)	90 (50%)	83 (45.9%)	0.4616
Had received at least one dose of PCV10, n (%)	92 (25%)	35 (19.4%)	56 (30.9%)	**0.0151**
Length of illness, median days (IQR)	4 (3-6) (n=3, missing data)	4 (3-6)	3 (2-5)	**0.0451**
Length of admission, median days (IQR)	6 (3-8) (n=8, missing data)	6 (4-10)	5 (3-7)	**<0.0001**
Blood white cell count, median $x10^9$/L (IQR)	12.7 (9.4-17.8) (n=3, missing data)	13.9 (9.6-19.4)	11.9 (9.2-15.8)	**0.0001**
Blood neutrophil count, median $x10^9$/L (IQR)	6.8 (4.2-11.7) (n=5, missing data)	8.6 (4.8-14.5)	6 (3.9-8.9)	**<0.0001**

4.3.1 Serotype distribution of nasopharyngeal samples analysed

117/368 (31.7%) patients had a pneumococcus isolated from their nasopharyngeal swabs and 113/117 were serotyped by standard microbiological techniques (figure 9). Serotype 14 (16/117, 13.7%), non-typeables (15/117, 12.8%) and serotype 6A (8/117, 6.8%) were the most prevalent serotypes. PCV10 serotypes accounted for 45/117 (38.5%) of the identified pneumococci.

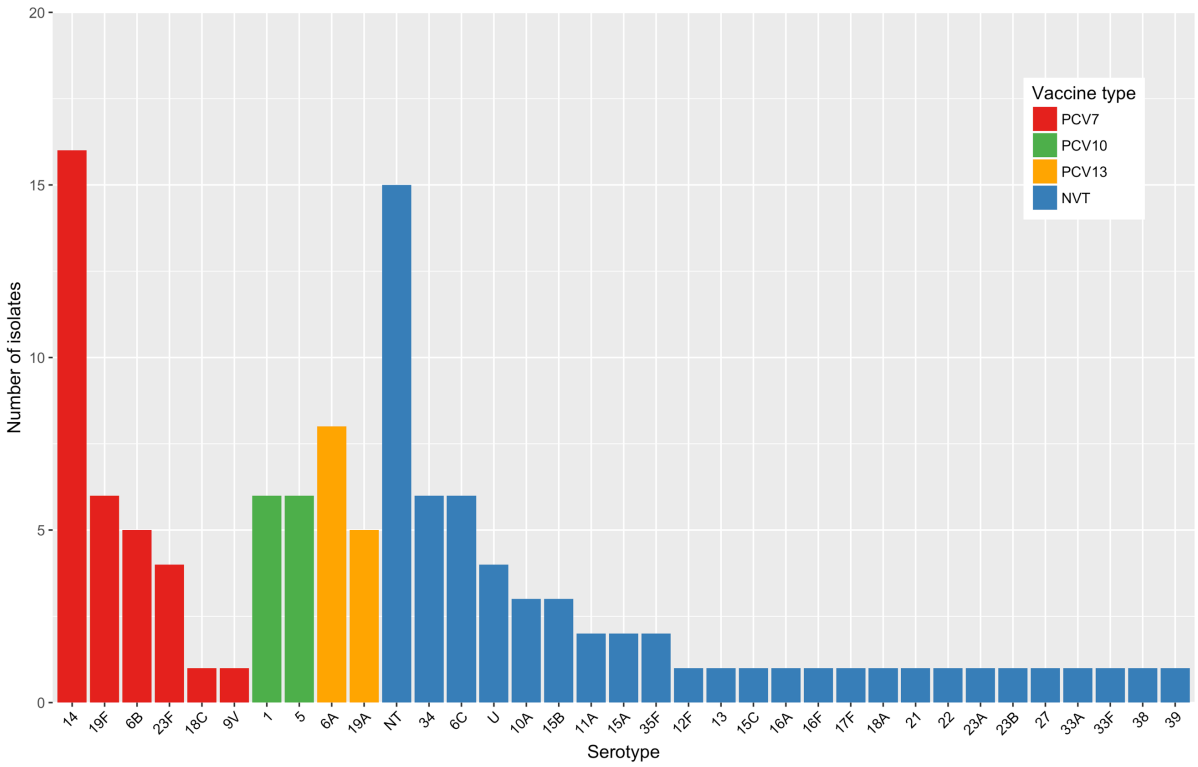


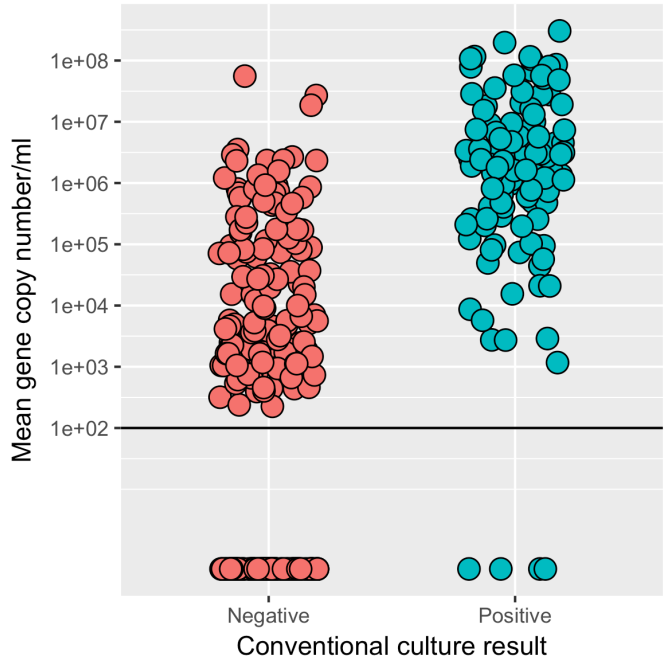
Figure 9. ***Serotype distribution of pneumococci identified by conventional culture from nasopharyngeal swabs collected from Nepalese children admitted with pneumonia.***
A total of 117/368 children had pneumococcus isolated by conventional culture and 113 of these had been serotyped. U=Not yet serotyped.

4.3.2 Comparison of conventional culture with *lytA* qPCR for detecting pneumococcus on nasopharyngeal swabs.

The detection of pneumococcus from nasopharyngeal swabs by conventional microbiology related well with a *lytA* copy number above the lower limit (10^2 copies/ml) of detection of the

qPCR assay (specificity 96.1%). However, there was a poor association between samples being negative by conventional culture and having a *lytA* copy number below the lower limit of assay detection (sensitivity 42.6%) (figure 10). The mean *lytA* copy number was significantly higher amongst carriers who had pneumococcus detected by conventional culture compared with those who did not ($10^{6.01}$ versus $10^{2.44}$ copies/ml, $p<0.0001$) (figure 10).

A



B

	qPCR +	qPCR -	Total
Culture +	**113**	**4**	**117**
Culture -	**152**	**99**	**251**
Total	**265**	**103**	**368**

Figure 10. ***Comparison of lytA copy number detected by real time PCR with conventional culture for the detection of pneumococcus from nasopharyngeal swabs collected from Nepalese children admitted with pneumonia.***

The mean *lytA* copy number was significantly higher amongst children who had pneumococcus detected by conventional culture compared with those that did not (A) ($10^{6.01}$ (95% CI $10^{5.72}$-$10^{6.31}$) versus $10^{2.44}$ (95% CI $10^{2.14}$-$10^{2.74}$) copies/ml, p<0.0001). Conventional culture had a specificity of 96.1% and sensitivity of 42.6% for detecting a *lytA* copy number greater than 10^2 (B).

4.3.3 Comparison of carriage density between children with pneumococcal pneumonia and children without pneumococcal pneumonia.

Pneumococcal pneumonia, determined by application of the composite diagnostic criteria, was detected in 18/179 (10.1%) of radiologically confirmed pneumonia cases (table 22). 5/18 were children who had blood cultures positive for pneumococcus. All children with a positive blood culture had end-point consolidation identified on their chest x-ray.

14/18 (77.7%) children with pneumococcal pneumonia had pneumococcus detected by conventional culture, compared with 46/161 (28.6%) of those children who had radiologically confirmed pneumonia but did not fit the criteria for pneumococcal pneumonia, and 54/181 (29.8%) of children admitted with pneumonia and had a normal chest x-ray (table 23).

Eight children were admitted with pneumonia but either did not have a chest x-ray performed or the chest x-ray was uninterpretable.

Table 22. ***Classification of pneumococcal pneumonia.***

Microbiological test	Abnormal chest x-ray and neutrophils $>7.5x10^9$	Abnormal chest x-ray only	Normal chest x-ray and neutrophils $>7.5x10^9$	Normal chest x-ray only
Serotype 1 carrier	**5**	**-**	**-**	**1**
Serotype 5 carrier	**2**	**1**	**2**	**1**
Serotype 14 carrier	**7**	**1**	**6**	**2**
***S. pneumoniae* detected on blood culture**	**5**	**-**	**-**	**-**

One participant had serotype 1 carriage and pneumococcus detected on blood culture.

Table 23. ***Pneumococcal carriage in Nepalese children admitted with pneumonia***

Characteristic	Pneumococcal pneumonia (PP) (N=18)	Non-pneumococcal pneumonia (NPP) (N=161)	Normal chest x-ray (N) (N=181)	p (PP vs NPP)	p (PP vs N)	p (NPP vs N)
Conventional culture positive	14/18 (77.8%)	46/161 (28.6%)	54/181 (29.8%)	**<0.0001**	**<0.0001**	0.8128
***lytA* PCR > 10^2 copies/ml**	18/18 (100%)	119/161 (73.9%)	122/181 (67.4 %)	**0.0080**	**0.002**	0.194
			***lytA* mean $\log_{10}$ copies/ml**			
Prior antibiotic therapy	6.5 (5.1-7.9)	2.7 (2.1-3.3)	2.8 (2.2-3.4)	**0.0004**	**0.0004**	0.8342
No prior antibiotic therapy	6.3 (5.7-6.8)	4.7 (4.2-5.2)	3.6 (3-4.2)	**<0.0001**	**<0.0001**	**0.0084**
Neutrophil count >$7.5x10^9$	6.3 (5.8-6.8)	3.8 (3.2-4.3)	3.8 (3.1-4.6)	**<0.0001**	**<0.0001**	0.8774

Non-pneumococcal pneumonia = children who do not fulfil the criteria for composite diagnostic pneumococcal pneumonia but were admitted with clinician diagnosed pneumonia and had an abnormal chest x-ray

4.3.4 Diagnostic accuracy of *lytA* qPCR for determining pneumococcal pneumonia.

The qPCR for *lytA* detected significantly higher gene copy numbers in those children with pneumococcal pneumonia compared with those who had non-pneumococcal pneumonia (abnormal CXR only) ($10^{6.3}$ mean copies/ml versus $10^{3.6}$ mean copies/ml, $p<0.0001$). The ROC curve for *lytA* qPCR showed good diagnostic accuracy (AUC=0.82) for distinguishing children meeting the definition of pneumococcal pneumonia from children with an abnormal chest x-ray who did not meet the diagnostic criteria for pneumococcal pneumonia (figure 11). Using a threshold of 568,921.2 copies/ml the *lytA* qPCR had a sensitivity of 0.72 and specificity of 0.72 when comparing children with pneumococcal pneumonia with those who did not meet the definition but had an abnormal x-ray.

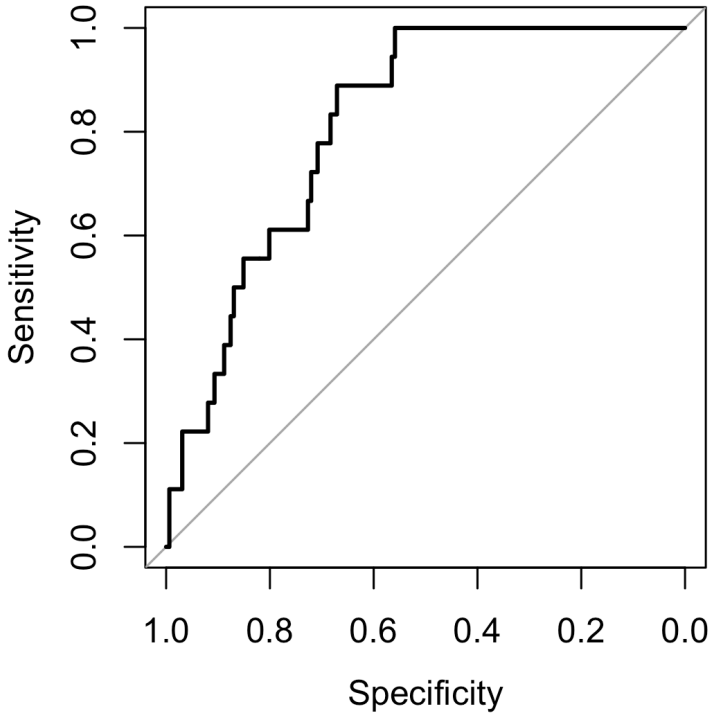


Figure 11. ***ROC curve demonstrating the performance of lytA copy number for differentiating Nepalese children admitted to Patan Hospital with defined pneumococcal pneumonia from those with who did not meet the criteria but had abnormal chest x-rays. AUC=0.82.***

The qPCR for *lytA* also detected significantly higher gene copy numbers in children with pneumococcal pneumonia compared with those who were admitted but had a normal CXR ($10^{6.3}$ mean copies/ml versus $10^{3.3}$ mean copies/ml, $p<0.0001$). The ROC curve showed a good diagnostic accuracy (AUC=0.82) for distinguishing children with pneumococcal pneumonia from children who were admitted with pneumonia but had a normal chest x-ray (figure 12). Using a threshold of 473,450.8 copies/ml the *lytA* qPCR had a sensitivity of 0.72 and specificity of 0.73 when comparing children with pneumococcal pneumonia with children who were admitted with pneumonia but had a normal chest x-ray.

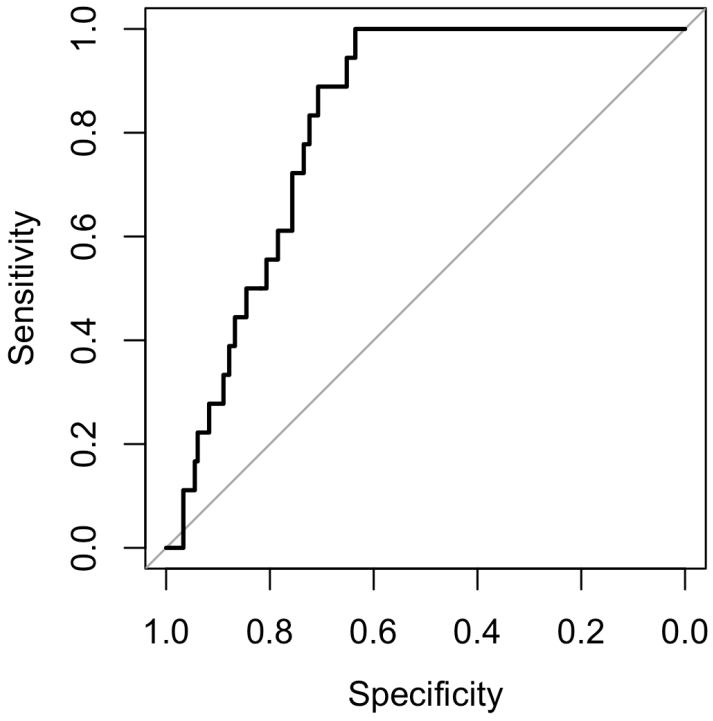


Figure 12. ***ROC curve demonstrating the performance of lytA copy number for differentiating pneumococcal pneumonia from non-pneumococcal non-radiologically confirmed pneumonia in Nepalese children admitted to Patan Hospital. AUC=0.82.***

4.3.5 Application of *lytA* copy number threshold value to the study cohort

Given the thresholds identified by analysis of the ROC curves, a threshold of 569,000 *lytA* copies was applied to those children in the study who had an abnormal chest x-ray. This showed

that a total of 59/179 (33%) children with abnormal chest x-rays would be classified as having pneumococcal pneumonia (figure 13). 42/59 (71.2%) of those children identified as having pneumococcal pneumonia using the threshold had end-point consolidation detected on their chest x-ray.

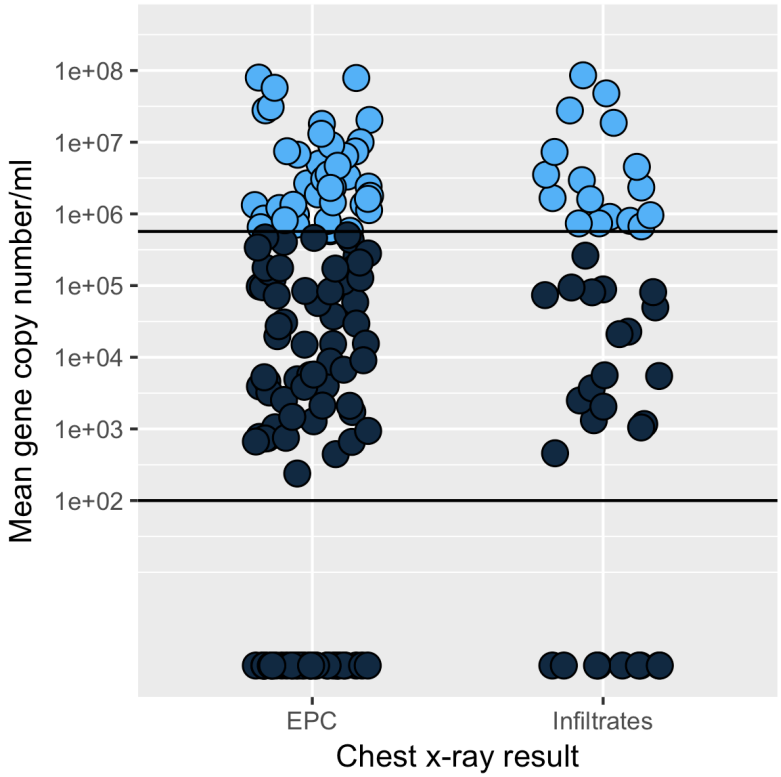


Figure 13. ***Detection of pneumococcal pneumonia in Nepalese children admitted to hospital with abnormal chest x-rays using a lytA copy number threshold of 569,000.***
Chest x-ray changes were classified as having end-point consolidation (EPC) or infiltrates. 42/136 children with EPC and 17/44 with infiltrates were classified as having pneumococcal pneumonia after applying the *lytA* threshold.

4.4 Discussion

This study gives evidence for the utility of *lytA* qPCR as a diagnostic test for pneumococcal pneumonia in children. Children admitted to hospital with pneumococcal pneumonia have mean *lytA* copy numbers $10^{2.7}$ and 10^{3} greater than children with non-pneumococcal radiologically confirmed pneumonia and non-pneumococcal non-radiologically confirmed pneumonia respectively. When a threshold of 569,000 *lytA* copies/ml was applied to the study cohort where radiological abnormalities were detected, 59/179 (33%) cases were identified as pneumococcal pneumonia compared with 18/179 (10.1%) using the composite diagnostic standard of; radiological changes, raised neutrophil count, and the presence of either serotype 1, 5, or 14 by conventional culture and serotyping.

4.4.1 *lytA* qPCR compared with conventional culture for detection of pneumococcus

This present study demonstrates that the *lytA* qPCR detects the presence of pneumococcus in many more samples when compared with conventional culture (72% versus 30.7%). This is broadly in keeping with one study in adults where *lytA* qPCR detected pneumococcus in 62.8% of participants compared with 44.9% using conventional culture.[59] This difference between qPCR and conventional culture is probably due to the fact that pneumococcus may not always grow on the agar plate during conventional culture and it may also only grow at a low density that would not be detected by the operator.

4.4.2 Accuracy of *lytA* qPCR for diagnosing pneumococcal pneumonia

This study shows that the density of pneumococcal carriage amongst Nepalese children, as determined by *lytA* qPCR, has a good ability at differentiating pneumococcal pneumonia from; children who do not meet the diagnostic criteria but who have chest x-ray changes, and children with non-pneumococcal pneumonia without chest x-ray changes. The sensitivity and specificity in this present study is in keeping with those described in studies conducted amongst

adults (see table 19). Notably in this present study the composite diagnostic used for classifying pneumococcal pneumonia was different from that applied in the adult studies. Further analysis of the clinical samples collected in this present study, for example detection of respiratory viruses on nasopharyngeal swabs and analysis of urine for pneumococcal antigens, could add further depth to the composite diagnostic framework in this study. The threshold values identified in the adult studies (between 8,000 and 100,000) were much lower than the one identified in this present study (569,000). This may be due to carriage density being higher in children than adults, or due to technical aspects around sampling and testing.

4.4.3 Strengths and weaknesses

The strengths of this study are the number of participants and the time-period over which samples were included. Those studies which have previously been conducted in adults have all had sample sizes less than 300. Whereas in this study samples were analysed from 368 participants to date and has the capability to further expand the sample size in the future. There is also the potential to meta-analyse the data in this current study with those described recently published multi-site study in children.[158] In this current study children were recruited across more than a two-year time period and thus the findings should not be strongly confounded by seasonal fluctuations in pneumococcal and non-pneumococcal causes of pneumonia.

Weaknesses of the study include; the number of pneumococcal pneumonia cases and the need for a chest x-ray to inform the diagnostic criteria. Although this study is larger than those studies conducted previously, further increasing the sample size would result in a greater number of pneumococcal pneumonia cases and thus would give greater accuracy to interpretation of this test using ROC curves. In this study radiological changes were used as part of the diagnostic criteria, however in many settings chest x-rays are not normal clinical practice for the management of pneumonia. Changes on the chest x-ray in this study were used to try and differentiate bacterial from viral aetiology. As such adding further criteria to the

diagnostic composite in this study, for example the detection of respiratory viruses and the presence of wheeze that is responsive to bronchodilators may be able to circumvent the need for using chest x-rays to differentiate participants.

4.4.4 Conclusions

This study shows that *lytA* qPCR is capable of detecting the presence of pneumococcus in a larger number of participants than conventional culture. In addition, measurement of pneumococcal carriage density in combination with radiological assessment of pneumonia has a good diagnostic accuracy for differentiating pneumococcal pneumonia from non-pneumococcal pneumonia in children admitted with clinician diagnosed pneumonia. Further work where the composite diagnostic used in this study is further developed, for example by adding additional assays which detect the presence of respiratory viruses, may be able to improve the diagnostic capability of *lytA* for pneumococcal pneumonia.

Chapter 5 The molecular epidemiology of pneumococci isolated from Nepalese children prior to PCV10 introduction

5.1 Introduction

Characterisation of pathogens at a molecular level and across a population provide an insight into the overall variation present within a species. Identification of population-wide variation is important because; it demonstrates the adaptability of the organism to selective pressures, the natural variation of the organism over time, and how any variation relates to quantifiable measures, such as disease causation.

These characteristics of an organism are important to identify because they allow us to understand the key biological processes of the organism, they provide opportunities to develop interventions (such as vaccines), and they allow us to anticipate how well an organism might respond to an intervention (for example, antibiotics).

5.1.1 Diversity and temporal changes across the pneumococcal genome prior to PCV use

It is well recognised that genetic variations within strains of pneumococcus are acquired over time, especially through the process of recombination. There are limited data however, on the natural variation of the whole pneumococcal genome over time within a population, as those studies which have been published to date have either collected isolates over a brief time period or had PCV introduction influencing strain diversity.[164,123] One study, of 127 isolates from Malawi collected over a five year period, used whole genome sequencing to show that different serotypes are comprised of different lineages.[125] A key limitation of this study however, is the relatively small sample size of isolates sequenced.

A much larger study in Maela, a refugee camp on the Thai-Burmese border, analysed 3085 carriage samples, although these were collected over a short time-period of 2 years, and showed

evidence for; capsular switching within lineages, the existence of common regions of recombination across lineages, and the role that NT pneumococci play in facilitating genetic recombination across the pneumococcal population.[124]

5.1.2 Genetic differences between carriage and disease isolates using a whole-genome approach.

It has long been observed that some pneumococcal strains are more associated with disease than asymptomatic carriage. As such it is important to examine what the genetic differences are between strains from each of these contexts, that is, pneumococci from healthy carriers compared with those found in disease. To date this has not been done at a whole genome level across a population. The above-mentioned study from Malawi reports findings from sequenced samples amongst both healthy carriers and patients with disease however, no comparisons were made between carriage and disease samples and such comparisons would be limited by the small sample size. One study compared the genomes of the same sequence type isolated from carriage with those isolated from disease, and in the two comparisons made, non-synonymous mutations were detected in the virulence factor peptidoglycan-*N*-acetylglucosamine deacetylase A (*pgdA*) of the carriage isolates.[165] Clearly limited conclusions can be drawn from the current data and more work examining differences in carriage and disease strains is needed.

5.1.3 Pneumococcal isolates collected at Patan Hospital provide a unique opportunity to explore the molecular epidemiology of pneumococcus

At Patan Hospital, Kathmandu, Nepal, there have been a series of studies which have been conducted since 2005. These studies have collected nasopharyngeal pneumococcal isolates from children who are healthy and residing in the community, nasopharyngeal pneumococcal isolates from children who were admitted with pneumonia, or isolates from children who had

a sterile-site culture positive for pneumococcus. All of these samples were collected prior to PCV introduction. As such the analysis of these isolates provides a unique opportunity to examine the molecular epidemiology of pneumococcus in this setting, outside of the influence of PCV selection.

5.1.4 Rationale for a study

A random selection of pneumococcal isolates from each of the above-mentioned study cohorts collected at Patan Hospital, Kathmandu, Nepal underwent whole-genome sequencing. The aim of this study is to examine the variation across the pneumococcal population in this setting prior to PCV introduction, and determine the differences between isolates collected from different contexts such as healthy carriage and disease.

5.2 Methods

5.2.1 Sample QC and filtering

All isolates were sequenced on the Wellcome Trust Sanger Institutes core sequencing pipeline. Sequenced isolates selected for analysis from Nepal were those which had a total yield less than 10,000 megebytes; less than 500 contigs; mapped to >60 % of *Streptococcus pneumoniae* ATCC 700669; had an assembly length between 1,900,000 - 2,300,000 bp; and greater than 20 times coverage. Selected isolates then underwent Kraken analysis for non-pneumococcal species contamination and any isolate with a content of more than 10% from another non-pneumococcal species was excluded.[166] Duplicate runs of the same isolate were also screened for and removed.

5.2.2 Coarse mapping and alignment creation

Sequence reads of the selected isolates were mapped against the reference genome of *Streptococcus pneumoniae* ATCC 700669 using SMALT.[167] Pseudogenomes generated using samtools mpileup were aligned using *Streptococcus pneumoniae* ATCC 700669 as the reference. SNPs were called across the alignment using SNP-sites.[168]

5.2.3 Maximum-likelihood phylogeny determination

The SNP alignment was used to generate a maximum-likelihood phylogeny using RAxML version 8.0.0 with 100 bootstrap replicates and gamma correction for among-site rate variation.[169] Phylogenies and associated meta-data were visualised using itol.embl.de.

5.2.4 Clustering using hierBAPS

The SNP alignment generated following coarse mapping to *Streptococcus pneumoniae* ATCC 700669 was used to estimate the population structure of the Nepal isolates using BAPS version 6.0.[170] By applying BAPS in a hierarchical manner a matrix of SNPs was generated for each

isolate. This matrix was then used to inform the 'pruning' of private mutations from the original SNP alignment using MATLAB version 2016a. The ensuing alignment was run again in BAPS to determine the final hierarchical population clustering profile.[171]

5.2.5 Estimating recombination and mutation rates and identification of recombination hotspots

Population clusters underwent fine mapping against either a previously published reference genome or a newly created custom reference (table 24). If there was no previously published reference which gave a mean mapping percentage of 90% or greater, then the isolate with the lowest number of contigs within a population cluster was chosen to be used as a custom reference. Custom reference assemblies were ordered against *Streptococcus pneumoniae* ATCC 700669 using Mauve.[172] Sequence reads were then mapped back onto these ordered assemblies using SMALT. Ordered assemblies were then visualised and curated using Artemis Release 16.[173]

Table 24. ***Details of references used for fine mapping of primary population clusters.***

Cluster name	Reference used	Serotype	Sequence type	Length, bp	Mean mapping percentage (range)	Accession number
BC6D	custom reference (lane 17138_5#59)	34	4560	2028491	95 (90.2-99.9)	-
BC23F-1	custom reference (lane 17138_5#19)	23F	7689	2076220	99.1 (98.2-99.7)	-
BC6A	custom reference (lane 17138_5#16)	6A	6031	2108999	97.1 (94.3-99.4)	-
BC19F	custom reference (lane 17138_5#15)	19F	3518	2073966	95.9 (94.4-99.2)	-
BC15B	Streptococcus_pneumoniae_15 BC_ST4209_v1.fa	15C	4209	1933597	95.7 (92.8-96.8)	-
BC35B	custom reference (lane 17138_5#37)	35B	8164	2089611	99.4 (97.8-99.7)	-
BC14	Streptococcus_pneumoniae_14_ ST63_v1.fa	14	63	2079395	95.6 (91.9-96.4)	CP001015
BC23F-2	custom reference (lane 17138_5#24)	23F	11011	2069612	98.7 (96.3-99.7)	-
BC1	Streptococcus_pneumoniae_P1 031_v1.fa	1	217	2111882	93.6 (91.4-94.7)	NC_012467
BC11A	custom reference (lane 16288_5#73)	11A	6025	2162630	96.8 (95.4-98.2)	-
BC10A	custom reference (lane 16288_5#64)	7B	1553	2106658	97 (95.1-99.2)	-

SNPs within sites of recombination were separated from SNPs outside of sites of recombination using Gubbins.[174] The data from this analysis were also used to determine the mean number of recombination sites to SNPs outside of recombination sites (r/m) ratio.

Recombination hotspots were classified as those regions across each population cluster which resided in the 95th centile or above for recombination site frequency. The sequences residing within the identified locations of recombination hotspots were extracted from the reference used for mapping each population cluster. These sequences were then compared with the described gene sequences of *Streptococcus pneumoniae* ATCC700669 using BLASTn comparison. The BLASTn hit with the highest Max score was the gene reported.

5.2.6 *In silico* PCR

In silico PCR was performed using an in-house script at the Wellcome Trust Sanger Institute which uses primer sequences to extract a desired DNA fragment from a contig or set of contigs. *In silico* PCR for categorising non-typeable (NT) pneumococci was conducted using primers as previously described.[175] The pspA and pspC loci were identified using primers previously reported.[176,177] pspA and pspC sequences were then aligned with Muscle V3.8.31 and phylogeny generated using RaxML version 8.0.0 with 100 bootstraps.[178]

5.2.7 Non-reference based assembly using Cortex

A non-reference based *de novo* assembly and genotyping of variations approach using coloured de Bruijin graphs by Cortex_var version 1.0.5.21. was applied to paired sequences from blood and CSF isolates.[179] The run_calls wrapper script of Cortex_var was used with stringent auto-cleaning, joint workflow, first k-mer length set at 31, and last k-mer length set at 61. Variants called by the run_calls wrapper script in the blood based isolate were then filtered according

to having passed the script filter and being greater than nine bases in length. A BLASTn search of variant sequences was then conducted against the whole *S. pneumoniae* ATCC 700669 genome. BLASTn matches were filtered based on an e-score <0.0001.

5.2.8 Typing the inverting variable region (*ivr*)

The *ivr* locus (figure 14) was characterised as previously described.[180] Briefly reads were mapped to the invariant region of the *Streptococcus pneumoniae* R6 spr0448 spnIVRhsdS CDS, and complementary read pairs which mapped with 95% or greater similarity to the N-terminus sequences (either segment A or B) were tallied. The N-terminus segment was then used as the reference sequence for determining the number of complimentary reads which mapped with 95% or greater similarity to the C-terminus segment sequences (either segment a, b, or c).

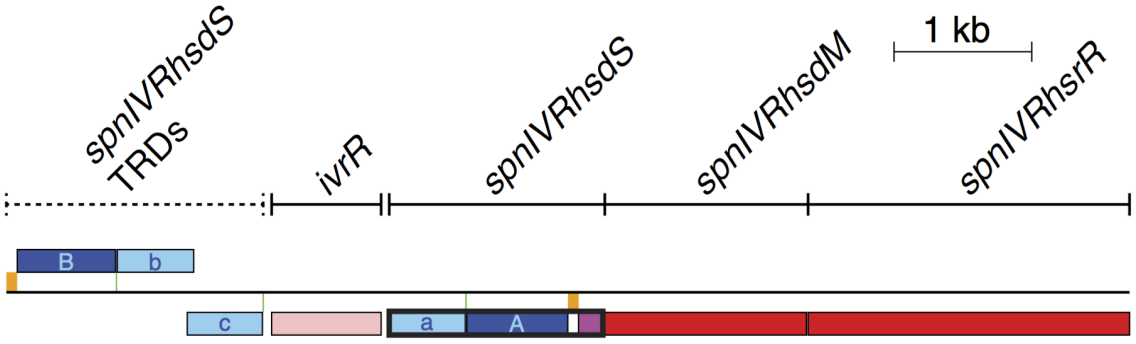


Figure 14. ***Structure of the Streptococcus pneumoniae R6 ivr locus as reported in figure 5 of Croucher et al (DOI: 10.1038/ncomms6471).***
Inversion of the genes within the locus results in an spnIVRhsdS allele which is either an A or B N-terminus combined with either an a, b, or c C-terminus.

5.2.9 *phtD* and *phtA* mapping and phylogeny creation

The sequences from the *Streptococcus pneumoniae* ATCC 700669 genome for *phtA* and *phtD* were used as references to map sequencing reads to using SMALT. Pseudogenomes were then generated and aligned as described above. The alignment was used to inform maximum likelihood phylogenies using RaxML as described above.

5.2.10 Statistical analysis

Proportions were compared by Fisher's exact test. All statistical analyses were performed using R version 3.3.1.[134]

5.3 Results

5.3.1 Population structure of pneumococci from children in Nepal prior to PCV10 introduction

Pneumococcal isolates were collected over a decade long period, between January 2005 and August 2014, from Nepalese children attending Patan Hospital, Kathmandu. This collection of samples consisted of a random selection from five studies (table 25). Three of these cohorts were nasopharyngeal carriage studies conducted among healthy children presenting to the paediatric outpatient's department. The fourth cohort was a nasopharyngeal carriage study of children admitted with pneumonia. The final cohort was those invasive disease isolates grown from sterile site cultures of children presenting to hospital (figure 15). Following whole genome sequencing, 458 isolates passed filtering and QC criteria for analysis.

Table 25. ***Overview of pneumococcal isolates sequenced and included in the analysis***

Cohort	Number of isolates, n (%)	Time period of collection	Age years, mean (range)	Male, %
HC 2009	107 (23.4)	February 2009 - April 2009	0.7 (0.1-1)	49.5
HC 2012	86 (18.8)	May 2012 - October 2012	0.7 (0.1-1)	46.5
HC 2014	125 (27.3)	April 2014 - October 2014	1.1 (0.4-4)	52
IP 2014	70 (15.3)	March 2014 - October 2014	2.1 (0-12.2)	61.4
IPD 2005-2014	70 (15.3)	January 2005 - July 2014	3.6 (0.2-14)	50

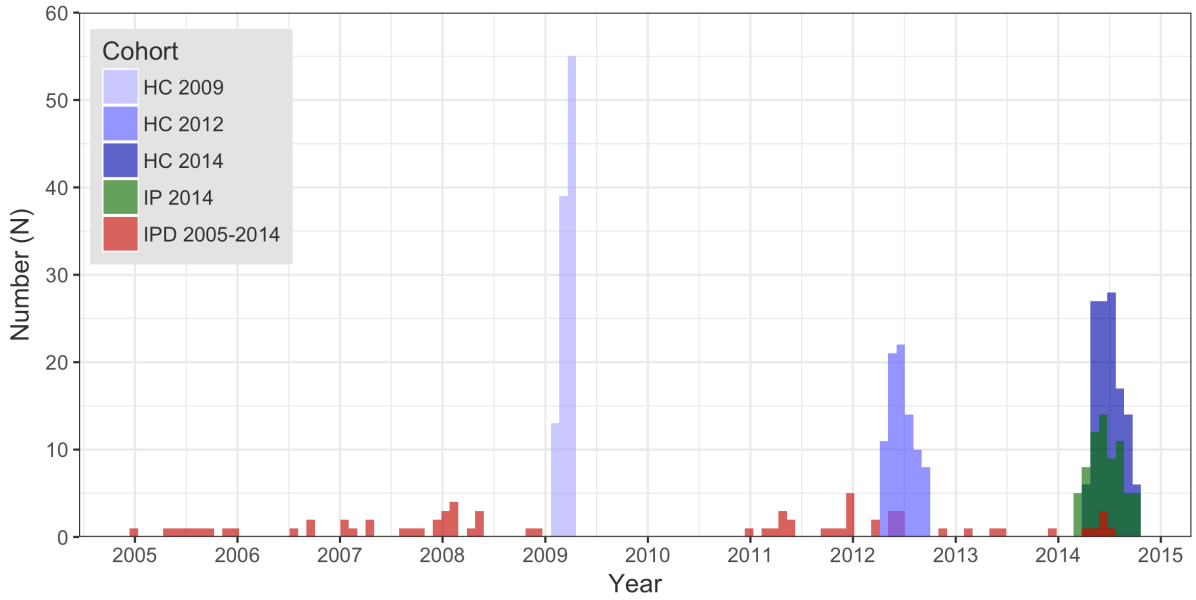


Figure 15. ***Number of sequenced pneumococcal isolates collected from Nepalese children at Patan Hospital, Kathmandu, Nepal between 2005 and 2014.***
Three cohorts were studies where isolates were collected from healthy children in three separate time-periods (HC 2009, HC 2012, and HC 2014). One cohort was a study where carriage samples were collected from children admitted to the hospital with pneumonia (IP 2014). In addition, isolates collected from sterile site cultures of children attending the hospital were analysed (IPD 2005-2014). PCV10 was introduced into the Kathmandu region in August 2015.

Whole genome sequences were mapped to the reference genome of *Streptococcus pneumoniae* ATCC 700669, SNPs identified and used to generate a maximum-likelihood phylogeny (figure 16). Descriptively the phylogeny appeared to have a structured population with discrete clades, however these clades appeared to have a balanced tree structure, evidenced by the many long terminal branches.

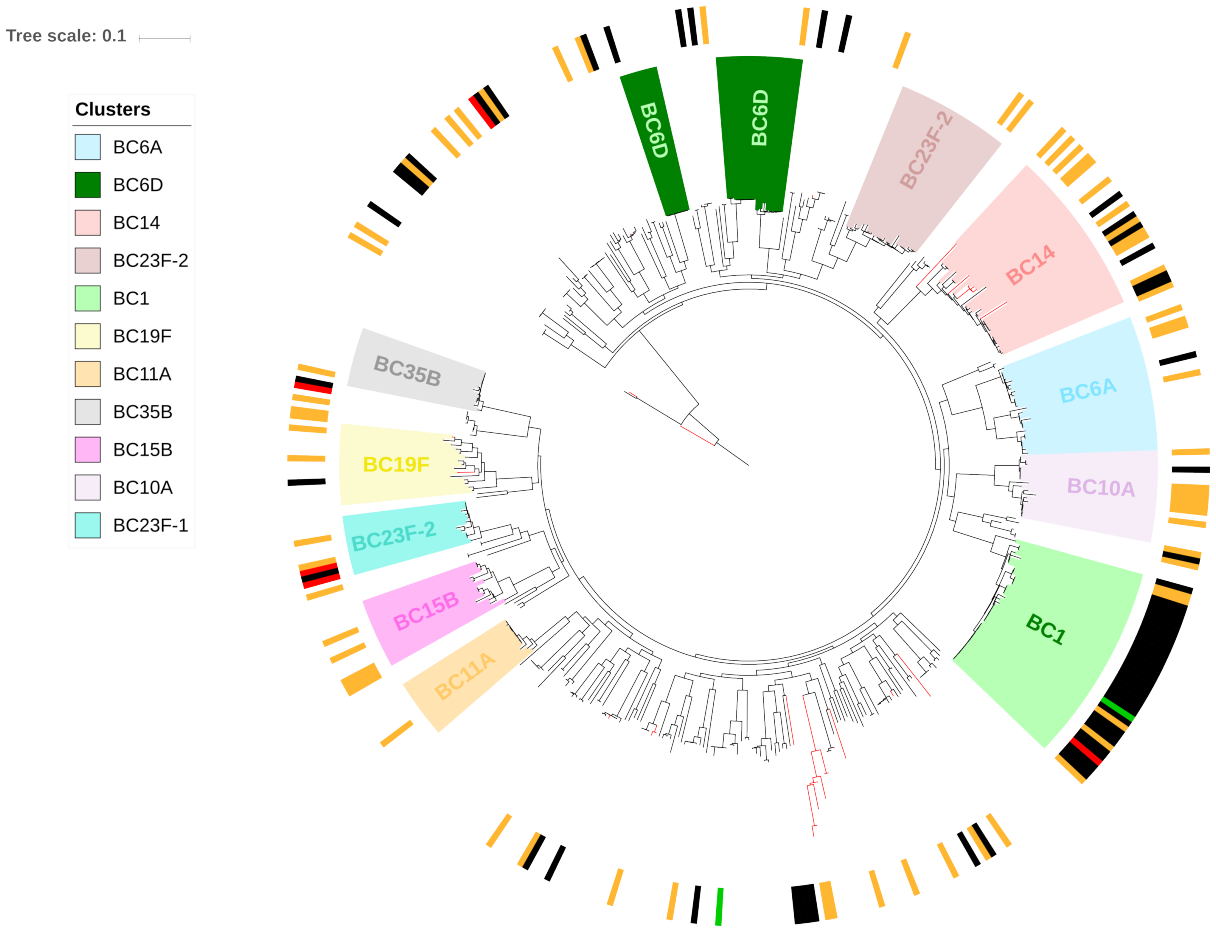


Figure 16. ***Single nucleotide polymorphism derived maximum-likelihood phylogeny of the pneumococcal population derived from Nepalese children prior to PCV introduction. The taxa belonging to the eleven largest primary clusters are highlighted.***
The outer ring indicates the source of each isolate, black=blood, red=CSF, green=other sterile site, orange=nasopharynx of child admitted with pneumonia, and white=healthy carrier. Nodes of non-typeable pneumococci are coloured red. The tree is rooted on the longest branch

5.3.1.1 Clustering of related sequences

Clustering was performed on the alignment of SNPs. Those mutations which were identified as being private were removed from the alignment so as to improve clustering accuracy. Thirteen monophyletic primary clusters were identified, representing 49.1% (225/458) of isolates (table 26). The largest monophyletic cluster was BC1 which consisted entirely of serotype 1 isolates found either in IPD or the nasopharynx of children with pneumonia.

Table 26. ***Characteristics of population clusters of pneumococci isolated from Nepalese children prior to PCV introduction.***

Cluster name	Number of isolates, n (%)	Serotypes present (n)
BC6D	23 (5)	6D (8), 34 (8), and 15A (7)
BC23F-1	11 (2.4)	23F (10), and NT (1)
BC6A	25 (5.5)	6A (12), 17F(6), 21(4), 9N (1), 19A (1), and NT (1)
bin	233 (50.9)	multiple
BC19F	15 (3.3)	19F (12), NT (2), and 19A (1)
BC15B	13 (2.8)	15B (8) and 15C (5)
BC35B	11 (2.4)	35B (10) and NT (1)
BC14	31 (6.8)	14 (25) and NT (6)
BC23F-2	21 (4.6)	23F (16), 19F (4), and 23B (1)
BC1	37 (8.1)	1 (37)
BC11A	11 (2.4)	11A (11)
BCNT-1	3 (0.7)	NT (2) and 17F (1)
BC10A	17 (3.7)	10A (7), 7B (6), 19F (3), and 19A (1)
BCNT-2	7 (1.5)	NT (7)

bin = isolates that could not be assigned to any other cluster

5.3.1.2 Serotypes within lineages and evidence for capsular switching

Eleven of the monophyletic clusters were represented amongst isolates from healthy carriers. Variations in the proportion of isolates that each cluster represented in each collection year were noted (figure 17). Most notable was the increase in representation of the BC14 cluster over time.

A total of 19 potential capsular switches were identified across the clusters which had capsule expressing pneumococci (table 27). Five of these identified switches included a switch to a non-typeable phenotype. BC6A is an example of where capsular switching within a lineage has occurred by recombination. Examination of the phylogeny indicates that there was a

serotype 17F and a serotype 21 lineage which are estimated to have diverged from each other in 1973. The serotype 21 lineage then further diverged into a serotype 6A lineage (figure 18).

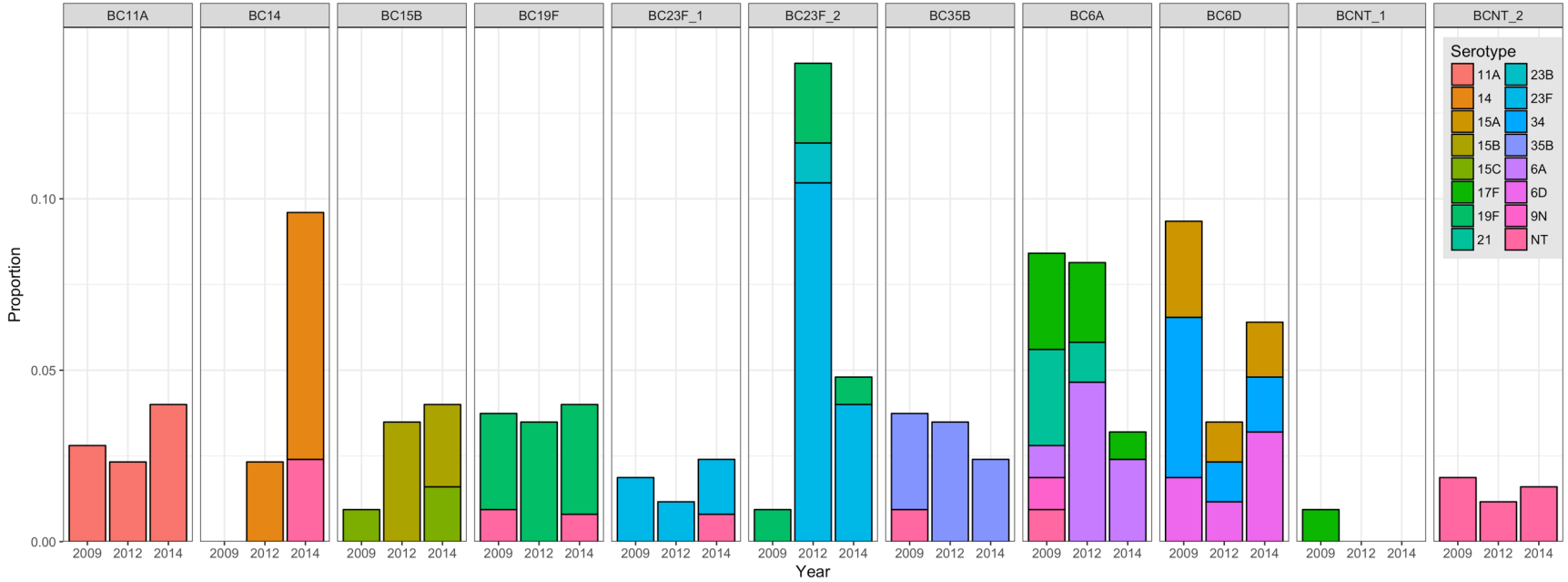


Figure 17. ***Proportion of total pneumococcal isolates sequenced in each collection year according to population cluster and serotype.***
All samples were isolated from nasopharyngeal swabs collected from healthy children visiting Patan Hospital, Kathmandu, Nepal prior to PCV introduction. Total number of isolates sequenced in each cohort were; 2009 N=107, 2012 N=86, and 2014 N=125.

Table 27. ***Capsular switches identified within each population cluster***

Population cluster	Number of different switches, n
BC6A	6 (which includes 1 NT)
BC10A	2
BC1	-
BC14	1 (which includes 1 NT)
BC6D	3
BC23F-2	2
BC11A	-
BC15B	1
BC23F-1	1 (which includes 1 NT)
BC19F	2 (which includes 1 NT)
BC35B	1 (which includes 1 NT)

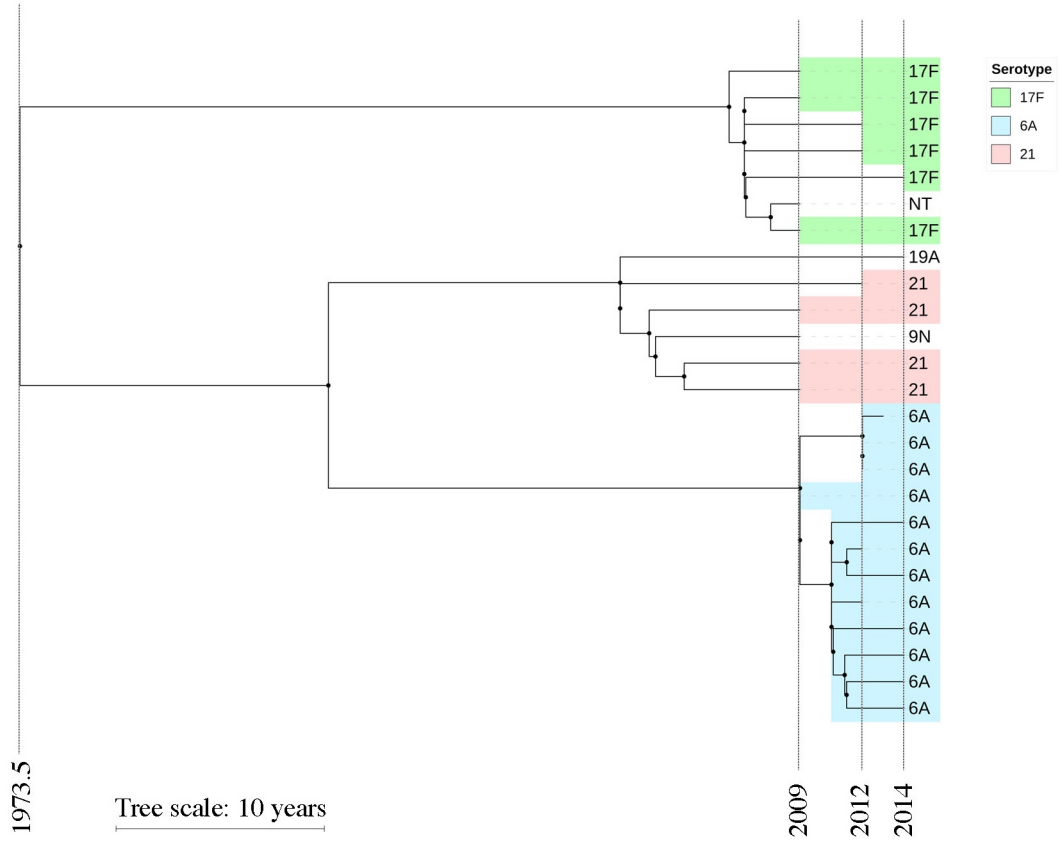


Figure 18. ***Time scaled SNP phylogeny created after removal of recombination regions for BC6A isolates collected from Nepalese children.***
This lineage diverged from being serotype 17F, to serotype 21, and then subsequently serotype 6A dominated. Taxa are coloured according to phenotypic serotypes. Phylogenetic dating using Path-o-gen estimates the most recent common ancestor to have occurred in 1973.

5.3.1.3 Non-typeable pneumococci

Non-typeable pneumococci have been shown to have a key role in the facilitation of genetic exchange across a pneumococcal population.[164] Given the large amount of diversity between and within lineages in this present study, the non-typeable pneumococci were further characterised.

There were 34 pneumococci which were non-typeable by serological methods. 11 of these isolates were associated with a capsule expressing lineage, the most prevalent of which was BC14. Non-typeable pneumococci were not represented by any dominant lineage, with 30 different sequence types represented across the 34 isolates. Categorisation of NTs based on the presence of previously described criteria,[175] demonstrates that the majority of NTs in this population possess *aliB* genes (table 28).

Table 28. ***Categories of non-typeable pneumococci from Nepalese children prior to PCV introduction.***

Category	Number of isolates detected, n (% of NTs)
Intact cps locus but not expressing	8 (23.5)
NT1 - cps deletion or partial deletion	1 (2.9)
NT2 - *nspA* sequence detected by in silico PCR	4 (11.8)
NT3 - *aliB1* detected	0 (0)
NT3 - *aliB2* detected	18 (52.9)
NT3 - *aliB1* and *aliB2* detected	3 (8.8)

*NB one isolate had both *nspA* and *aliB2* detected

5.3.1.4 Recombination and mutation rates across population clusters

Regions of recombination and SNPs outside of these regions were determined. For each population cluster the mean number of recombination events to mutation events (r/m) was calculated, and showed that BC1 had the lowest mean r/m ratio (0.04, 95% CI 0.01-0.06) and BC19F had the highest mean ratio (0.23, 95% CI 0.16-0.29) (figure 19). BC14 is comprised of both serotype 14 and NT isolates. Comparing the r/m ratios between the two serotypes within

BC14 showed that NT isolates had a significantly higher mean r/m ratio (0.27 vs 0.12, p=0.0386).

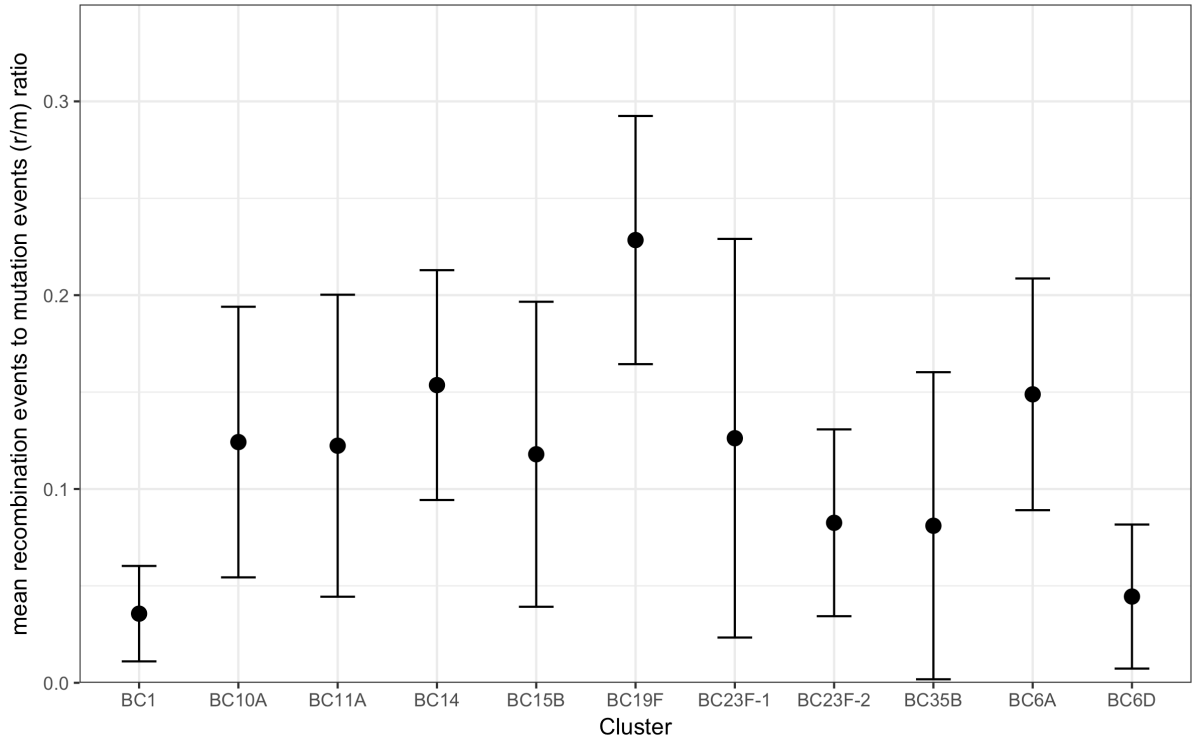


Figure 19. ***Mean recombination events to mutation events (r/m) in each pneumococcal population cluster and 95% CIs.***
Recombination and mutation events were determined for each isolate in each population cluster and the arithmetic mean calculated.

5.3.2 Recombination hotspots

Hotspots of recombination for each population cluster were classified as those sites with a recombination frequency equal to or above the 95th centile. Genes within these sites were identified by comparing sequences within hotspots with described genes in the *Streptococcus pneumoniae* ATCC 700669 genome. This yielded 281 described genes to be present in a hotspot in more than one of the population clusters. The genes which were in hotspots of recombination across the most clusters were *pspA*, SPN23F03020, SPN23F03030, *mraW*, *ftsL*, *pbp2x*, *arcA*, and *pspC* (table 29). Interestingly the two surface proteins *pspA* and *pspC* were

identified as hotspots in the same clusters. Similarly, SPN23F03020, SPN23F03030, *mraW*, *ftsL*, and *pbp2x* were identified in hotspots from identical clusters.

Table 29. ***Described genes most frequently detected in population cluster hotspots of recombination***

CDA	Product	Hotspot clusters
pspA	pneumococcal surface protein A	BC10A, BC14, BC15B, BC19F, BC23F-1
SPN23F03020	LacI family regulatory protein	BC6A, BC14, BC15B, BC23F-1, BC23F-2
SPN23F03030	membrane protein	BC6A, BC14, BC15B, BC23F-1, BC23F-2
mraW	S-adenosyl-methyltransferase MraW	BC6A, BC14, BC15B, BC23F-1, BC23F-2
ftsL	cell division protein	BC6A, BC14, BC15B, BC23F-1, BC23F-2
pbp2x	penicillin binding protein 2x	BC6A, BC14, BC15B, BC23F-1, BC23F-2
arcA	arginine deiminase	BC1, BC6A, BC6D, BC15B, BC19F
pspC	pneumococcal surface protein C	BC10A, BC14, BC15B, BC19F, BC23F-1

5.3.2.1 Characterisation of pneumococcal surface protein A

In silico PCR detection of *pspA* across the population demonstrated that the vast majority 433/458 (94.5%) of isolates possessed a *pspA*. Only 1/70 IPD isolates, 1/70 inpatient carriage isolates, and 23/318 healthy carriage isolates did not have a *pspA* gene detected. The phylogeny of the detected genes demonstrated the previously described six clades which reside across three distinct families (figure 20).[181] Clades 1 and 2 reside within family 1, with clade 1 (161/433) representing the majority of isolates compared with clade 2 (53/433). Within family 2 clades 3, 4, and 5 each represented 101, 73, and 30 isolates respectively. The vast majority of IPD isolates (47/69, 68.1%) had a pspA from clade 1. Fifteen isolates had a *pspA* from clade 6, the majority of which (14/15) were of a NT phenotype. Three NT isolates appeared to be mistyped when examining the *pspA* phylogeny, one was classified as clade 6 but resides within family 1 on the phylogeny, one was classified as clade 2 but appears to be an outgroup or more closely related to family 3, and another isolate was classified to clade 1 but appears to be more related to family 2. Each of these three isolates had BLASTn bit scores much lower than the

majority of the other isolates, thus indicating the BLASTn alignment was likely to be unreliable in these few circumstances.

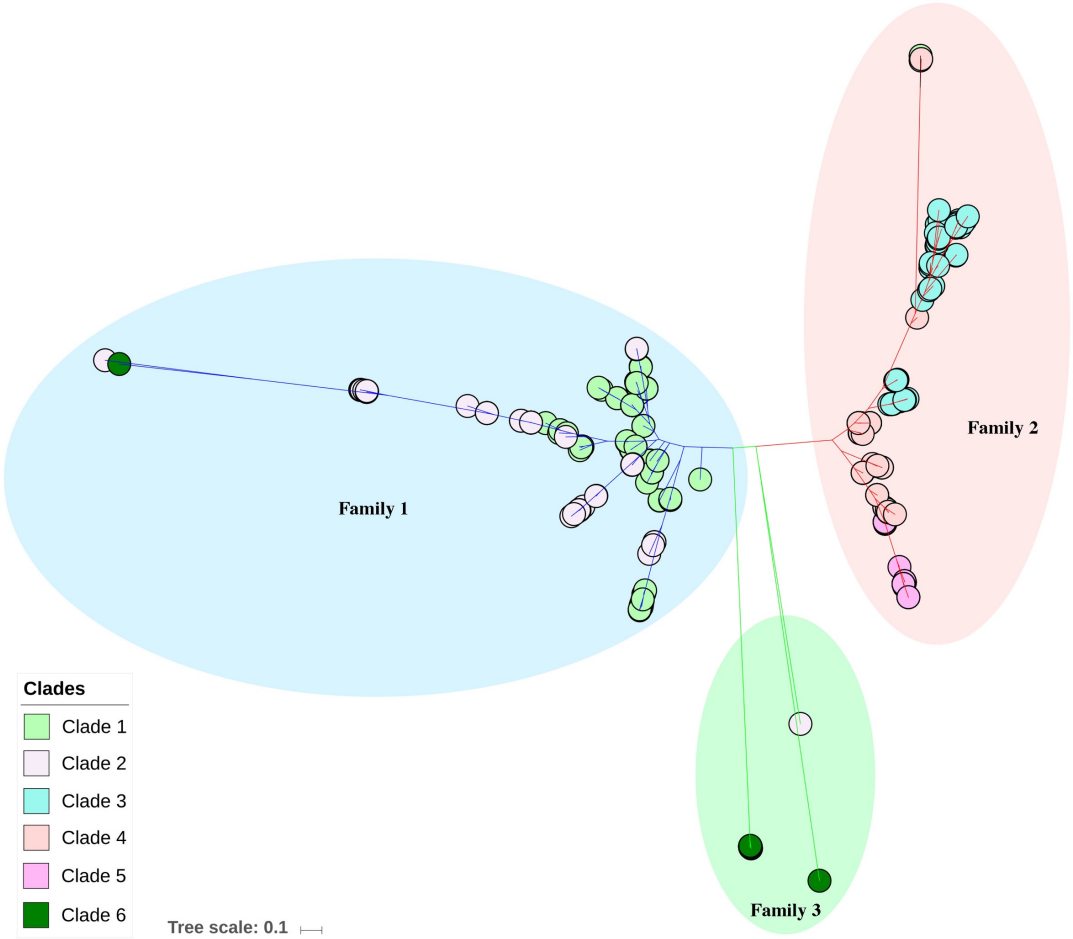


Figure 20. ***Unrooted phylogeny of identified pspA sequences using in silico PCR of pneumococcal whole genome sequences sourced from samples collected from Nepalese children prior to PCV introduction.***
Clades were determined by BLASTn comparison with previously described *pspA* sequences.

5.3.2.2 Characterisation of pneumococcal surface protein C

pspC was identified across the sequenced population by *in silico* PCR. This demonstrated 252/458 (55%) isolates to have a *pspC* locus detected using previously described definitions. *pspC* was more frequently found in IPD isolates 56/70 (80%) compared with inpatient carriers 41/70 (58.6%), and healthy carriers 155/318 (48.7%) (IPD vs inpatients p=0.0099, IPD vs

healthy carriers p<0.0001). There was no significant difference when comparing the proportion of inpatient carriers that had a *pspC* detected with healthy carriers (inpatients vs healthy carriers p=0.1479). The observed differences between disease and carriage isolates appeared to be related to the fact that both serotype 1 (BC1) and 14 (BC14) lineages are highly represented in disease isolates (figure 21). BLASTn comparison with previously described *pspC* alleles demonstrates the majority of identified *pspC* alleles belong to the choline anchored group (244/252, 96.8%) as opposed to the LPTXG anchored group (8/252). Of the IPD isolates with a detected *pspC* allele the majority were found to be pspC2.2 (20/56, 35.7%) and pspC6.13 (12/56, 21.4%) (table 30).

Within population clusters, for example BC1 and BC14, there appeared to be a dominant sub-type of *pspC*. However, it is interesting to note that even within BC1, overall a very conserved lineage with a low r/m ratio, variation in the sub-type of *pspC* was detected (figure 22).

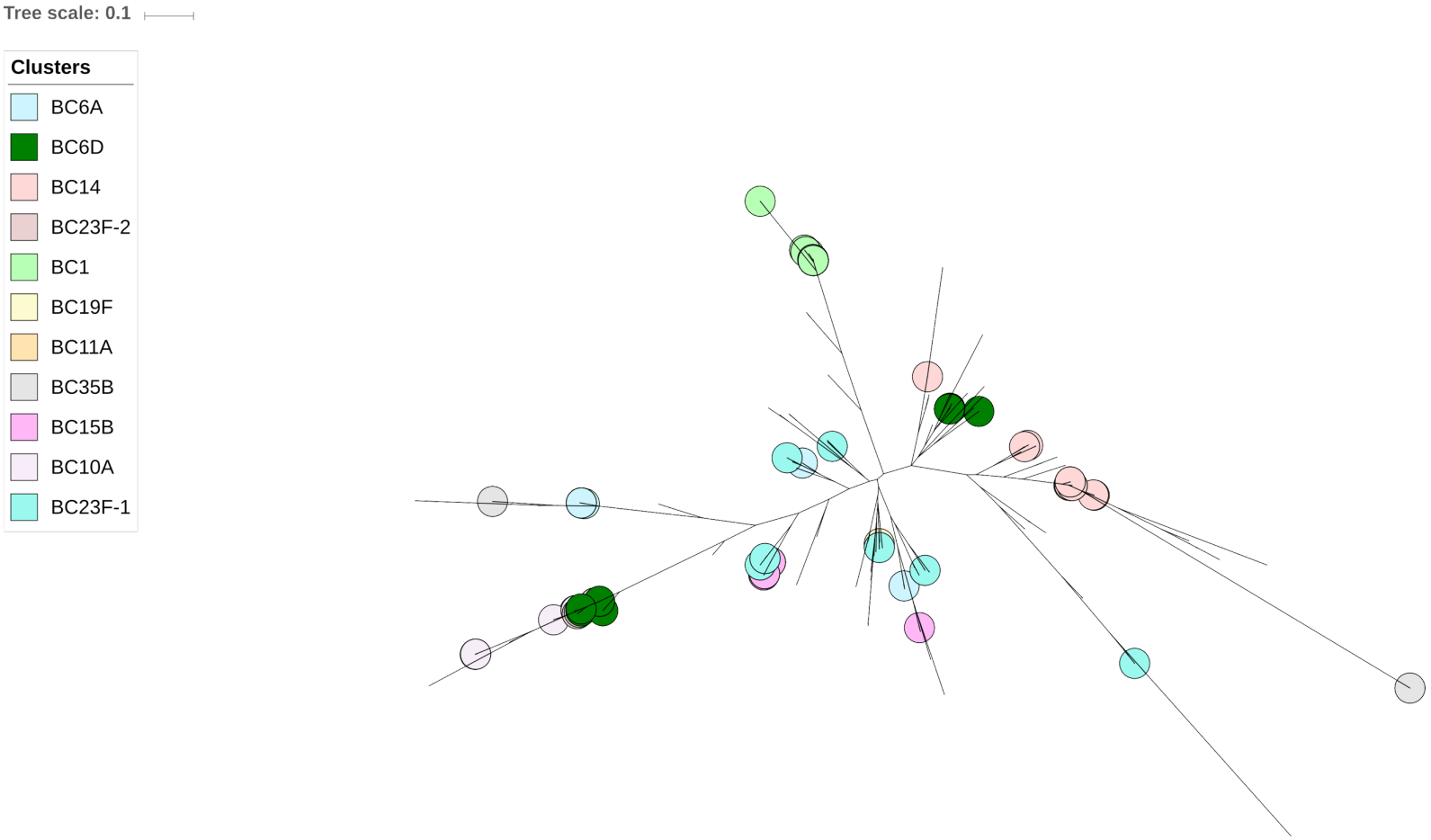


Figure 21. ***Unrooted maximum likelihood phylogeny of the 252 pneumococcal surface protein C (pspC) alleles detected by in silico PCR across the population of 458 pneumococcal isolates from Nepalese children.***
Previously described primers defining the *pspC* locus were used to conduct *in silico* PCR of pneumococcal whole-genome sequences for alleles of *pspC*. Sequences for these alleles were then aligned and a maximum likelihood phylogeny generated using this alignment. Isolates which belong to a population cluster, defined by clustering similar sequences from the entire population of 458 pneumococcal whole genomes are coloured accordingly. Uncoloured taxa were not assigned to one of the monophyletic clusters.

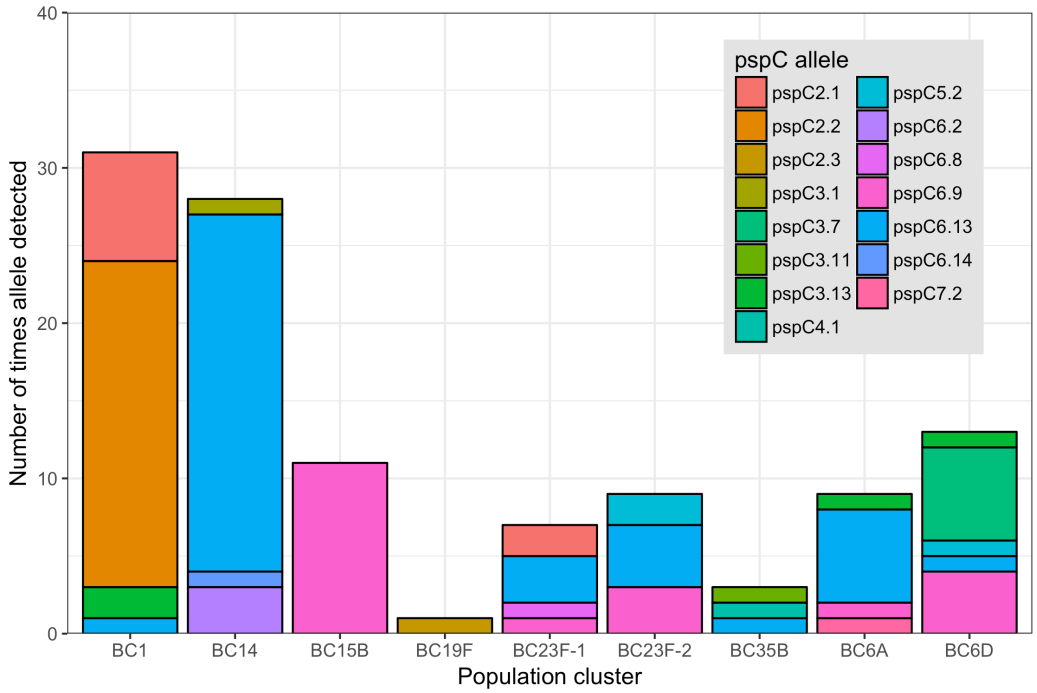


Figure 22. ***Presence of pspC sub-type alleles according to pneumococcal population cluster.***
In silico PCR was used to identify *pspC* alleles and these sequences were then compared with previously described alleles using BLASTn. All sub-types, other than pspC7.2, in this visualisation are choline anchored pspCs.

Table 30. ***pspC alleles detected within IPD isolates.***

pspC allele	Population cluster	Frequency, n
pspC2.1	BC1	7
pspC2.2	BC1	20
pspC2.3	BC19F	1
pspC2.5	bin	2
pspC3.1	bin	1
pspC3.13	bin	3
pspC3.13	BC1	2
pspC3.7	BC6D	1
pspC3.8	bin	1
pspC4.1	bin	1
pspC5.1	bin	1
pspC5.2	BC10A	1
pspC6.13	BC14	6
pspC6.13	bin	4
pspC6.13	BC1	1
pspC6.13	BC23F-1	1
pspC6.8	BC23F-1	1
pspC6.9	BC23F-1	1
pspC6.9	bin	1

bin = cluster to which all isolates are assigned if unable to be assigned to any other cluster

5.3.3 Comparing isolates collected in the same patient taken from blood with those taken from CSF.

Three children had paired blood and CSF isolates sequenced. These pairs of sequences were each a serotype 8, 9V, and 23F. Following comparison of blood with CSF sequences, variations of 10 bp or more in length underwent BLASTn comparison with the gene coding regions of *Streptococcus pneumoniae* ATCC 700669. Five described genes were identified to be varied between the blood and CSF sequence in all three pairs (table 31).

Table 31. ***Genes which were identified to be varied between all three pairs of blood and CSF sequences.***

CDS	product
SPN23F04610	type I restriction-modification system S protein (representative of the hsdS allele of the *ivr* locus)
SPN23F05800	putative sugar phosphotransferase system (PTS),IIA component
SPN23F09290	phtD pneumococcal histidine triad protein D
SPN23F10770	phpA putative streptococcal histidine triad protein
SPN23F10980	sugar phosphotransferase system (PTS), IIA component

5.3.3.1 The inverting variable region (*ivr*) locus

Prior work has shown that the *ivr* locus (a type I restriction modification system) has a role in phase variation and that phases may vary according to the environment, for example the nasopharynx or blood, that the pneumococcus reside within.[32] As such the *ivr* locus was further characterised across the sequenced Nepalese pneumococcal population. The *ivr* locus consists of an N-terminus segment and a C-terminus segment. Two different types of N-termini components (A and B) and three different types of C-termini components (a, b, and c) have been described.

Characterisation of the *ivr* locus demonstrated 247 isolates to have a dominant A allele and 192 isolates to have a dominant B allele (figure 23). Nineteen isolates had ten or less read pairs mapped to the *ivr* alleles. 83 isolates appeared to possess only either an A or a B locus. To confirm that *ivr* alleles were not detected because of assembly errors, the number of sequenced

reads which mapped directly to the N-terminus of the *ivr* locus was assessed, and showed that there were very low numbers of reads mapping to those isolates which did not have an *ivr* locus detected by the initial read pair assessment.

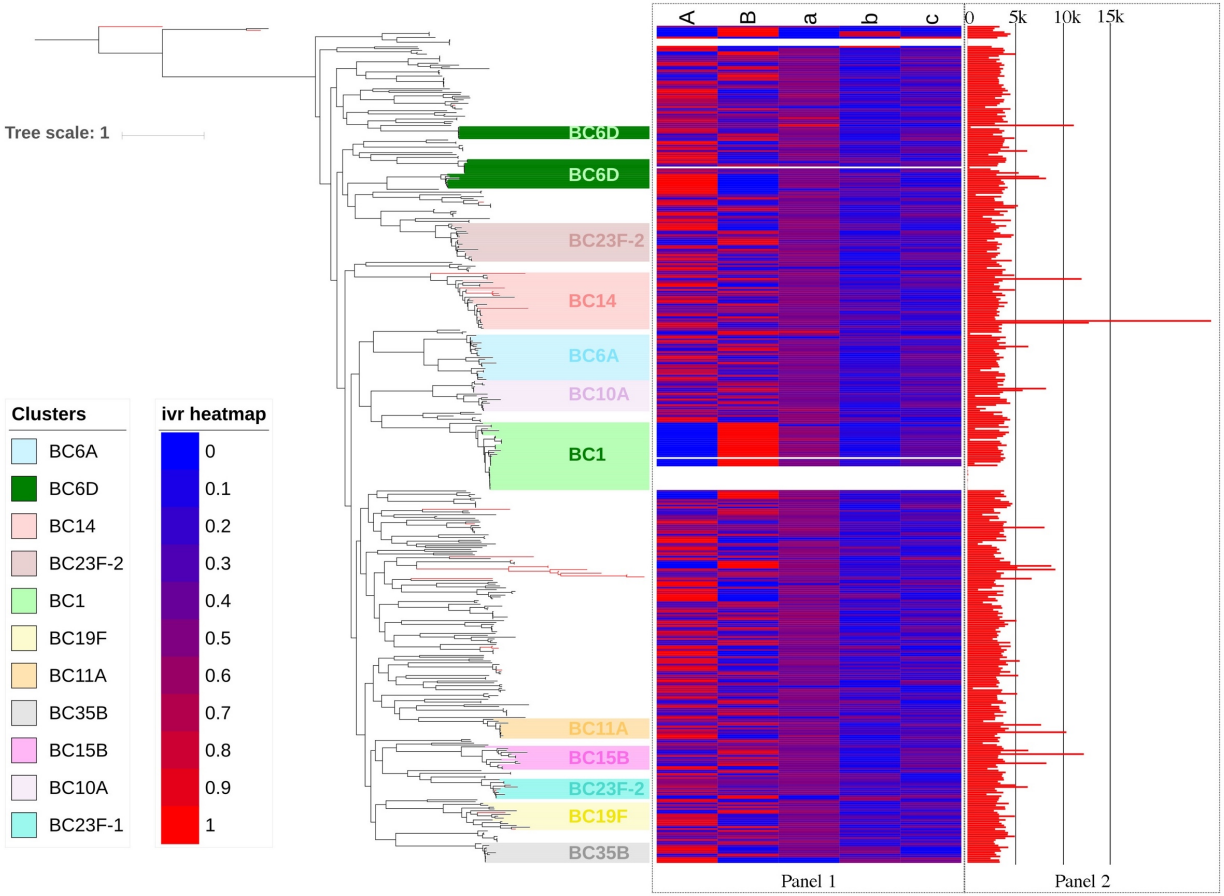


Figure 23. ***Single nucleotide polymorphism derived maximum-likelihood phylogeny of the pneumococcal population amongst Nepalese children prior to PCV introduction. The taxa belonging to the eleven largest primary clusters are highlighted.***
Panel 1: Heatmap demonstrating the proportion of read pairs mapped to each allele of the *ivr* locus of sequenced pneumococcal isolates. The A and B columns represent the proportion of reads detected across the N-terminus portion of the *ivr* locus. The a, b, and c columns represent the proportion of reads detected across the C-terminus portion of the *ivr* locus. Panel 2: The bar graph demonstrates the number of reads that were directly mapped to the N-terminus regions of the *ivr* locus

5.3.3.2 Pneumococcal histidine triad proteins A and D

Pneumococcal histidine triad proteins are surface bound proteins which have been described as having a role in epithelial adhesion. As such both pneumococcal histidine triad protein A (*phtA)* and pneumococcal histidine triad protein D (*phtD)* were further characterised across the study population.

Sequenced reads were mapped to the *phtA* gene from the *Streptococcus pneumoniae* ATCC 700669 genome and a phylogeny generated (figure 24). At least 90% coverage of the reference *phtA* gene was detected in 66/70 (94.3%), 67/70 (95.7%), and 279/318 (87.7%) of isolates in the IPD, inpatient carriage, and healthy carriage cohorts respectively. Interestingly all serotype 3 isolates had less than 90% coverage of the *phtA* gene.

Mapping of reads was also performed against the *phtD* gene from the *Streptococcus pneumoniae* ATCC 700669 genome and a phylogeny created (figure 25). At least 90% coverage of the reference *phtD* gene was detected in 63/70 (90%), 63/70 (90%), and 294/318 (92.5%) of isolates in the IPD, inpatient carriage, and healthy carriage cohorts respectively.

For both *phtA* and *phtD*, within primary population clusters including those population clusters which appeared to be very clonal such as BC1, there appeared to be wide variability in the gene sequences identified.

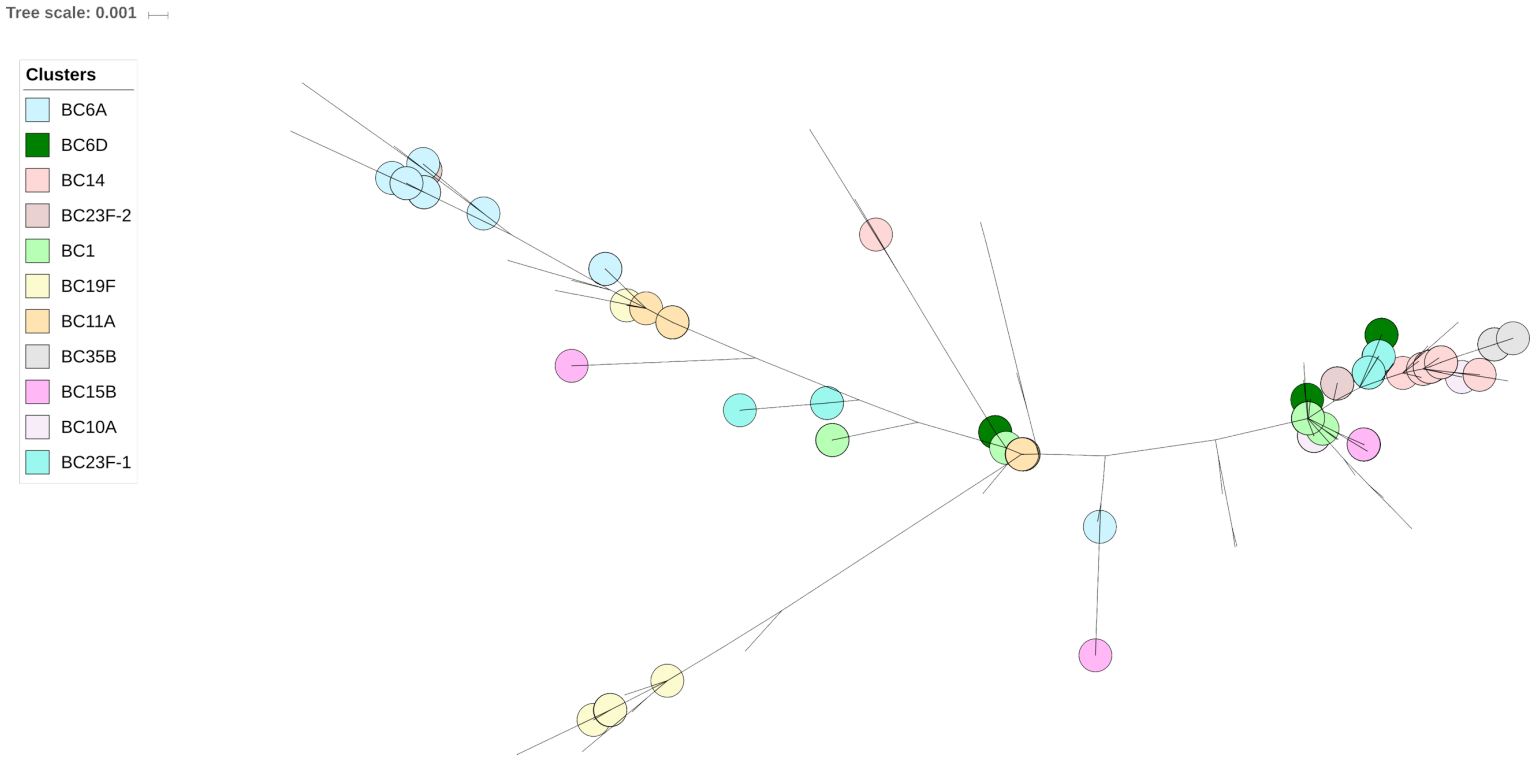


Figure 24. ***Unrooted maximum likelihood phylogeny of pneumococcal histidine triad protein A (phtA) in pneumococcal isolates from Nepalese children.***
Whole genome sequences of pneumococci isolated from Nepalese children were mapped against the *phtA* of the *Streptococcus pneumoniae* ATCC 700669 genome. Mapped sequences were then aligned and a maximum likelihood phylogeny generated from this alignment. Taxa belonging to the primary population clusters defined by clustering similar pneumococcal whole genome sequences across the entire 458 isolates have been coloured according to the included legend.

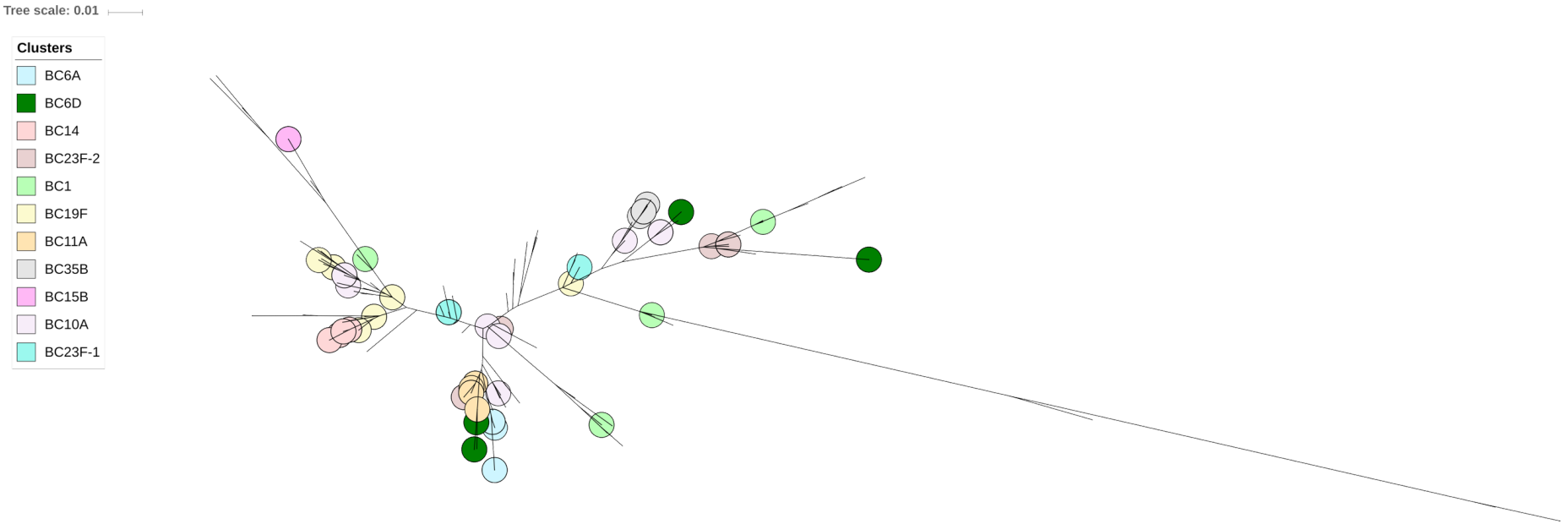


Figure 25. ***Unrooted maximum likelihood phylogeny of pneumococcal histidine triad protein D (phtD) in pneumococcal isolates from Nepalese children.***
Whole genome sequences of pneumococci isolated from Nepalese children were mapped against the *phtD* of the *Streptococcus pneumoniae* ATCC 700669 genome. Mapped sequences were then aligned and a maximum likelihood phylogeny generated from this alignment. Taxa belonging to the primary population clusters defined by clustering similar pneumococcal whole genome sequences across the entire 458 isolates have been coloured according to the included legend.

5.4 Discussion

This study demonstrates the plasticity of the pneumococcal genome across the study population and within specific lineages prior to PCV introduction in Nepal. Host immunity and antibiotic use appear to influence this genetic plasticity, as evidenced by the number of population cluster recombination hotspots that genes such as *pspA*, *pspC*, and *pbp2x* were found within. It was also evident from this study that capsular switching occurs naturally outside of the influence of PCV use. Finally, comparison of isolate pairs from blood and CSF indicates that there are variations in the *ivr* locus, which are known to be associated with phase variation, and also variations in the genes encoding the surface proteins *phtA* and *phtD*, both of which are associated with the adhesion of pneumococci to host cells.

5.4.1 Capsular switching events

Theoretically capsular switching may be more likely in some lineages than in others. This is not only because each lineage has a different intrinsic rate of recombination and mutation, but also because the mechanism for capsular switching is more simple in some cases than others. For example, the single nucleotide mutation required to switch from serotype 11A to 11D, whereas the recombination of an entire *cps* locus was required to switch from serotype 21 to 6A.

The total number of capsular switching events identified in this study population (19 episodes across a population 458 isolates) appears to be consistent with a prior study conducted in Maela (191 episodes across a population of 3085 isolates).[164] This would suggest that the frequency of capsular switching is generally consistent across populations. It has been known for over a century that the host immune response to capsule polysaccharide can confer subsequent protection. As such the evolutionary role of capsular switching is likely to be one of escape from host immune selection.

5.4.2 Non-typeable pneumococci in Nepal arise sporadically.

When comparing healthy carriers in this study with the Maela study population, there was a significantly smaller proportion of NT pneumococci noted (23/318, 7.2%, in Nepalese healthy carriers versus 512/3085, 16.6%, in Maela, $p<0.0001$). This is likely to be a result of two factors. Firstly, the Maela study population was dominated by a sporadic NT lineage with 128/512 NTs in this population belonging to this group. Secondly, none of the NTs detected in this current study had a sequence type consistent with the classic NT lineages previously described, whereas the Maela study detected 44/512 NTs that belonged to the classic sequence type 448.[182] This indicates that NTs arise sporadically in this current study population. An observation which is further evidenced by the fact that 30 different sequence types are represented across the 34 identified NTs.

5.4.3 Recombination rates and recombination hotspots

In general, the range of mean r/m ratios appeared to be consistent with those described in the Maela population. Although BC1 and BC6D isolates were very clonal and had r/m ratios lower than any population cluster described in the Maela study. The low r/m ratios did not appear to be related to disease with the BC6D lineage primarily identified in healthy carriers (21/23 isolates), whilst the BC1 lineage was only identified in IPD (32/37) and inpatient carriage (5/37) samples.

There was some similarity between the present study and the Maela study, in the genes identified in recombination hotspots. Most notable of these were *pspA*, *pspC*, and *pbp2x*.

Both *pspA* and *pspC* have been described as having roles in attenuating components of the human innate immune system. Thus it is not surprising that these genes are under selective pressure. Beta-lactam antibiotics were being used in increasing amounts across the duration of

the current study and would explain the observed selective pressure on pneumococcal penicillin binding proteins.

5.4.4 Variations found between blood and CSF pairs and the use of these findings to explore variation across the population

The exploration of variation between isolates collected from the same patient but different sites in this present study provided the opportunity to identify genes that may have key functions in allowing pneumococcus to enter and survive in different host environments.

This present study identified the *ivr* locus as varying between the blood and CSF states. Given that prior work has shown this locus to vary between nasopharyngeal and blood environments in animals it seems likely that this finding is real and biologically plausible.[32] A prior study has also shown that different combinations of *ivr* alleles correlate with different patterns of DNA methylation and it is this pattern of methylation which confers the ability of the bacteria to exchange genetic material.[180] Interestingly in the present study some isolates appeared to either lack a detectable *ivr* locus or lacked the ability to vary it. Of most note was the finding that all serotype 1 isolates belonging to BC1 either only possessed a B N-terminus allele or lacked a detectable N-terminus allele all together. This would suggest that lineages such as this have limited capacity for genetic uptake and may explain the low r/m detected.

The pneumococcal histidine triad proteins have previously been described as having a role in host cell adhesion.[183] As such it is plausible that these proteins do not have a necessary function in the blood stream but may have a role in traversing the blood brain barrier. This role may explain why variation is seen between blood and CSF isolates, as the detected variation may be the reflection of molecular processes that result in genetic down-regulation.

A prior study has also addressed the blood compared with CSF question in a similar manner using 674 pairs of pneumococcal isolates collected in the Netherlands. This study showed there

was no clear evidence for adaptation of pneumococcus between the blood and CSF environment.[184] Although it should be noted that SNPs within genes were focussed upon in the Dutch study, whilst the present study assessed larger sized variations, which would most likely be due to recombination and/or rearrangement of alleles.

5.4.5 Strengths and weaknesses.

Limitations of the study are the small number of samples included in some lineages and cohorts, and the single site within which samples were collected. A greater number of samples analysed may have allowed a better impression of the inter-lineage dynamics and increased further the accuracy of recombination hotspot detection within a lineage. Widening of the geographical area would allow more certainty in how well the observations made in this study could be generalised to other settings in Nepal.

Strengths of the study include the long time period over which IPD samples were collected, the multiple time periods over which healthy carriage samples were collected, and the absence of PCV influence on the sample population. The time periods over which the samples were collected means that there is greater confidence in observations made around lineage changes when compared to studies which use short cross-sectional observational periods. PCV introduction has been shown to be a strong influencer of population dynamics and as such the absence of the effect in this population provides the opportunity to focus on the other environmental influences at play.

5.4.6 Conclusions

This study demonstrates that the pneumococcal population in Nepal varies widely over time, with recombination playing a key role in the manifestation of this variation. Regions of recombination do not appear to be random with hotspots of recombination observed to be common across different population clusters. Interestingly the most common hotspots of

recombination detected reflected selective pressures, such as the host immune system (for example *pspA* and *pspC*) and antibiotic use (for example *pbp2x*). A comparative genomic approach to a small number of paired isolates collected from blood and CSF demonstrates variations in genes, such as the *ivr* locus and the pneumococcal histidine triad proteins, that are biological relevant to environmental change. Characterisation of the *ivr* locus across the present studies population showed that some lineages have either a restricted set of alleles or do not have any detectable alleles at all. A finding which may indicate that certain lineages are more capable of invading into the blood stream but not as successful at crossing the blood brain barrier. Confirmation of these findings in larger data sets is needed.

Chapter 6 Antibiotic resistance patterns of pneumococcus in Nepal with a molecular context.

6.1 Introduction

Access to antibiotics in resource-limited settings has increased dramatically over the last decade. Unfortunately, legislation governing use, and the enactment upon such legislation, has generally lagged behind, leading to rapidly increasing, largely unrestricted, availability of antimicrobials in settings such as Nepal. This is exemplified by one small study in Kathmandu, Nepal which showed street-side pharmacies dispense medications with minimal medical history taking or consideration for legal restrictions.[111] Such widespread unregulated use of antibiotics can result in the rapid development of antibiotic resistance leading to the drugs becoming clinically redundant and poor treatment outcomes due to treatment failures.

6.1.1 Antibiotic use in Kathmandu, Nepal

There are no published data on patterns of antibiotic use in Nepal, however one paper reports an analysis of antibiotic consumption between 2000 and 2010 for India and China.[112] These data should be reflective of Nepali trends in antibiotic consumption as India and China are Nepal's main trade partners. These data show that in the 2000 to 2010 period global antibiotic consumption increased by 36%. Most notably in 2010, India and China were the world's first and second largest consumers of antibiotics respectively. In India and China, the three classes of antibiotics that were most consumed were beta-lactams, macrolides, and fluoroquinolones. Due to the lack of published data, more specific information on antibiotic usage in Kathmandu, Nepal was collected through interviews with two senior paediatricians who have been working at Patan Hospital, Kathmandu for the last two decades (table 32). These data showed that both beta-lactams and co-trimoxazole have been the mainstay of treatment for children with pneumonia for the last 10 years. It is also important to note that enteric fever has a high incidence in Nepal, and use of macrolides and fluoroquinolones to treat it is prevalent.[185]

Table 32. ***Paediatric antibiotic usage patterns at Patan Hospital, Kathmandu, Nepal.***

	Fluoroquinolones
Pre 2005-2015	Used for urinary tract infections and dysentery. Also used as second line treatment for enteric fever following azithromycin.
Pre 2005-2011	Ofloxacin used for outpatient managed enteric fever.
	Beta-lactams
Pre 2005-2015	Ampicillin used for children admitted with pneumonia who have had no prior antibiotic treatment.
Pre 2005-2015	Ceftriaxone used as second line inpatient treatment of pneumonia or if the patient has had prior antibiotics in the community.
Pre 2005-2015	Amoxicillin commonly used in the community for pneumonia over the last two decades
2010-2015	Cefixime also used for UTIs, enteric fever, and dysentery
2010-2015	Cefixime commonly used in community for pneumonia
	Co-trimoxazole
Pre 2005-2015	Used for UTI prophylaxis and PCP pneumonia.
Pre 2005-2014	Widely used in the community for pneumonia until 2014
	Macrolides
2013-2015	Azithromycin used for outpatient managed enteric fever (clinically diagnosed).
Pre 2005-2015	Azithromycin occasionally used as a single agent for pneumonia.
Pre 2005-2015	Azithromycin occasionally used in the community for pneumonia (consistent for last two decades)
	Tetracyclines
Pre 2005-2015	Occasionally used in older children for cholera and scrub typhus
	Chloramphenicol
Pre 2005-2015	Occasionally used for children admitted with pneumonia. Rarely used and/or available in the community.

6.1.2 Pneumococcal antibiotic resistance in Nepal

There are limited data on pneumococcal antibiotic resistance patterns in Nepal and those data which are available are from the assessment of samples collected no later than 2007 (table 33). In general, these studies indicate that pneumococcal resistance to co-trimoxazole is high, whilst resistance to other antimicrobial classes was within ranges that would not affect clinical practice around using these drugs. The high levels of co-trimoxazole resistance is consistent with the reported long-term and widespread use of this antibiotic until 2014 for community acquired pneumonia. Notably since 2014, amoxicillin (in place of co-trimoxazole) has been used as first-line treatment of children with community acquired pneumonia. Thus the lack of current data on pneumococcal antibiotic susceptibilities and the subsequent changes in guidance on antibiotic use underlie the need for further assessment of pneumococcal antibiotic susceptibility in this setting.

Table 33. ***Studies which have reported pneumococcal antibiotic sensitivities in Nepal***

Study (publication year)	Region	Sample type	Time period of collection	Samples, n	Findings
Joshi *et al* (2008)[117]	Children's homeless shelter in Ranibari, Kathmandu, Nepal	Nasopharyngeal swabs	not reported	185	• 0% non-susceptible to penicillin
Shah *et al* (2009)[119]	Children presenting to Kanti Children's Hospital, Kathmandu, Nepal	Blood and CSF	Nov 2004 - Mar 2007	28	• 4% non-susceptible to penicillin • 68% non-susceptible to co-trimoxazole • 7% non-susceptible to erythromycin • 4% non-susceptible to cefotaxime • 0% non-susceptible to chloramphenicol
Khanal *et al* (2010)[116]	Children and adults presenting to BP Koirala Institute of Health Sciences, Dharan, during	Blood, CSF, pleural fluid, and corneal ulcers	Apr 2005 - Apr 2006	26	• 0% non-susceptible to penicillin • 23.1% non-susceptible to co-trimoxazole • 0% non-susceptible to erythromycin • 0% non-susceptible to cefotaxime • 0% non-susceptible to chloramphenicol • 4% non-susceptible to ciprofloxacin • 0% non-susceptible to vancomycin
Easow *et al* (2011)[118]	Children and adults presenting to Manipal Teaching Hospital, Pokhara	Sputum, bronchoalveolar lavage, blood, and ocular specimens	Jan 2000 - Aug 2007	312	• 5% non-susceptible to penicillin • 34.3% non-susceptible to co-trimoxazole • 7.4% non-susceptible to erythromycin • 11.1% non-susceptible to chloramphenicol • 0.4% non-susceptible to tetracycline

6.1.3 Genetic determinants of antibiotic resistance

Antibiotic use places selective pressure upon the pneumococcal population. Pneumococci can attain the ability to resist the effect of antibiotics by either; acquisition of genes which inhibit or restrict the antibiotics from reaching the target site of action, or by varying the intended target site of the antibiotic through the acquisition of genetic variations.

Prevention of antibiotics reaching the intended target site can take place through a variety of mechanisms. Notably the genes encoding the proteins which carry out these actions are typically contained within mobile genetic elements enabling easy transmission between bacteria.

Preventing the antibiotic from entering the bacterial cell through the regulation of cell wall porins has been described for Gram-negative bacteria, but is yet to be described for pneumococcus.[186] In a similar manner the removal of antibiotics from the bacterial cell can be expedited through energy dependent efflux pumps, an example of which is *mefA* which encodes a macrolide efflux pump that has been detected in pneumococci.[187] If the antibiotic does make it into the bacterial cell, the target site can still be protected through modification, for example the erythromycin ribosome methylase *ermB* methylates the 23S rRNA protecting the binding site from erythromycin.[188] In addition some bacteria can acquire enzymes which act to inactivate the antibiotic directly, for example beta-lactamases which act to break the beta-lactam ring structure of beta-lactam antibiotics.[189]

Antibiotics typically bind the intended target protein with very high affinity. If a gene encoding the target protein however, was to acquire a point mutation which leads to reduced binding affinity, then strains containing this gene would proliferate. Additionally, target proteins can be altered through recombination, with the mosaicism of pneumococcal penicillin binding proteins, through recombination with closely related *Streptococcus spp.*, a prime example of this process.[190]

6.1.4 Rationale for the study

Therefore, this study was conducted with the aim of determining the antibiotic resistance patterns of the pneumococci amongst Nepalese children, including the examination of how these patterns change over time, and between carriage and disease isolates. In addition, it was aimed to identify the main molecular characteristics that relate to pneumococcal antibiotic resistance, for each of the main antibiotic classes being used in Nepal.

6.2 Methods

6.2.1 Study design

All of the isolates described in table 25 of chapter 5, which underwent whole genome sequencing, were processed, as described below, for antibiotic sensitivities using disc diffusion. Antibiotics tested were chloramphenicol, erythromycin, tetracycline, co-trimoxazole, oxacillin, and norfloxacin. The primary objective of the study was to determine the patterns of antibiotic susceptibility across the pneumococcal population amongst Nepalese children prior to PCV introduction. The secondary objectives were to determine the molecular variations that relate to antibiotic non-susceptibility in this population for each tested antibiotic.

6.2.2 Determining pneumococcal antibiotic sensitivities.

Pneumococcal antibiotic sensitivities were assessed according to European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines. Frozen pneumococcal isolates were defrosted and plated on Columbian agar with 5% sheep blood and incubated in 5% carbon dioxide overnight at 37.5°C. Colonies were then emulsified in 3 ml of peptone water until turbidity was consistent when compared with a McFarland 0.5 standard. A bacterial lawn was then created by dipping a swab into the suspension and streaking it across an Iso-sensitest (Oxoid, Basingstoke, United Kingdom) with 5% horse blood agar plate. Antibiotic discs (Oxoid, Basingstoke, United Kingdom) were then placed on the plate before incubating in 5% carbon dioxide overnight at 37.5 °C. Zone edges were read from the front of the plate with the lid removed and in reflected light. The EUCAST breakpoint table version 6.0 (valid from the 1st of January 2016) was used to interpret susceptibility zone diameters (table 34).

Table 34. ***Antibiotic disc concentrations and respective breakpoints taken from the EUCAST breakpoint table.***

Antibiotic disc	Disc concentration (μg)	Sensitive if diameter greater than or equal to (mm)	Resistant if less than (mm)
Chloramphenicol	30	21	21
Erythromycin	15	22	19
Tetracycline	30	25	22
Trimethoprim-sulfamethoxazole[1]	1.25-23.75	18	15
Oxacillin (benzylpenicillin breakpoint)	1	20	20
Oxacillin (amoxicillin breakpoint)	1	8	8
Norfloxacin	10	12	12

[1]Breakpoints are expressed as the trimethoprim concentration**.**

6.2.3 Detection of antibiotic resistance genes

SRST2 short read sequence typing using the Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT) database of resistance genes as a reference was applied to the paired reads of whole genome sequences from each isolate.[191,192] SRST2 was used with the default settings. Specifically, a minimum coverage cut-off of 90%, maximum divergence of 10%, and maximum number of mismatches per read of 10. The ARG-ANNOT database accessed on the 4th of November 2015 was used.

6.2.4 Mapping of sequence reads to reference alleles and generation of phylogenies.

The sequences from the *Streptococcus pneumoniae* ATCC 700669 genome for *pbp1a*, *pbp2a*, *pbp2x*, *folA*, *folP*, *gyrA*, and *parC*, were used as references to map sequencing reads to using SMALT. Pseudogenes were then generated and aligned as described in section 5.2.2. The alignment was used to inform maximum likelihood phylogenies using RaxML as described in section 5.2.3.

6.2.5 Statistical analysis

Proportions were compared by Fisher's exact test. All statistical analyses were performed using R version 3.3.1.[134] Phylogenetic visualisations were created using iTOL (www.itole.de).[193]

6.3 Results

450 of the 458 pneumococcal isolates which underwent whole-genome sequencing had antibiotic sensitivities to chloramphenicol, erythromycin, tetracycline, trimethoprim-sulfamethoxazole (co-trimoxazole), oxacillin, and norfloxacin assessed. Overall 359/450 (79.8%) of isolates were non-susceptible to at least one of the antibiotics tested (figure 26).

Whole genome sequences were screened for antibiotic resistance genes contained within the ARG-ANNOT gene database using short-read-sequence typing. A total of 246 genes within 359 of the 450 isolates were detected. The most prevalent gene detected was *tetM* 155/246 (63%). The presence of a resistance gene that confers resistance to a given antibiotic class was compared with the detection of non-susceptibility/susceptibility to chloramphenicol, erythromycin, and tetracycline.

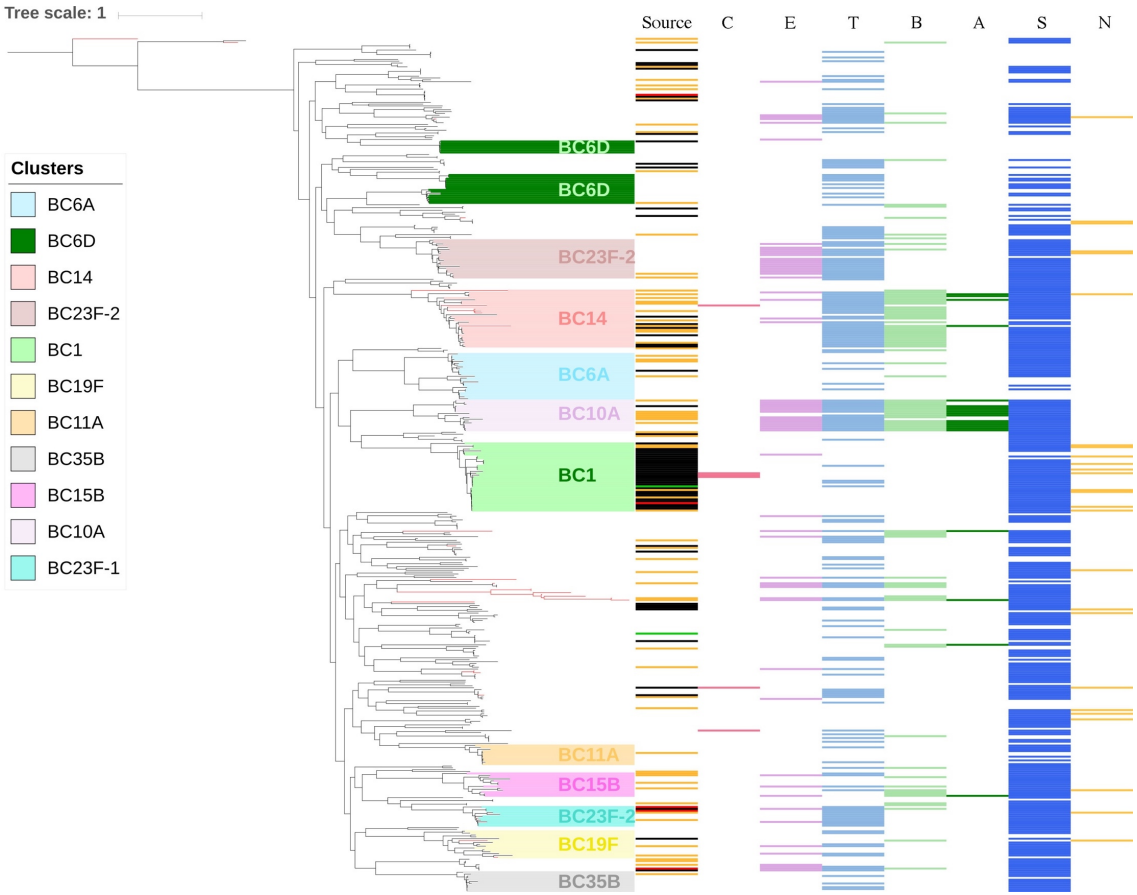


Figure 26. ***Single nucleotide polymorphism derived maximum-likelihood phylogeny of the pneumococcal population derived from Nepalese children prior to PCV introduction.***

The taxa belonging to the eleven largest primary clusters are highlighted. The tree is rooted on the longest branch. The 'Source' column indicates the source of each isolate, black=blood, red=CSF, green=other sterile site, orange=nasopharynx of child admitted with pneumonia, and white=healthy carrier. Coloured bars; in column 'C' indicate chloramphenicol resistant isolates, in column 'E' indicate erythromycin resistant isolates, in column 'T' indicate tetracycline resistant isolates, in column 'B' indicate benzylpenicillin resistant isolates, in column 'A' indicate amoxicillin resistant isolates, in column 'S' indicate co-trimoxazole resistant isolates, and in column 'N' indicate norfloxacin resistant isolates.

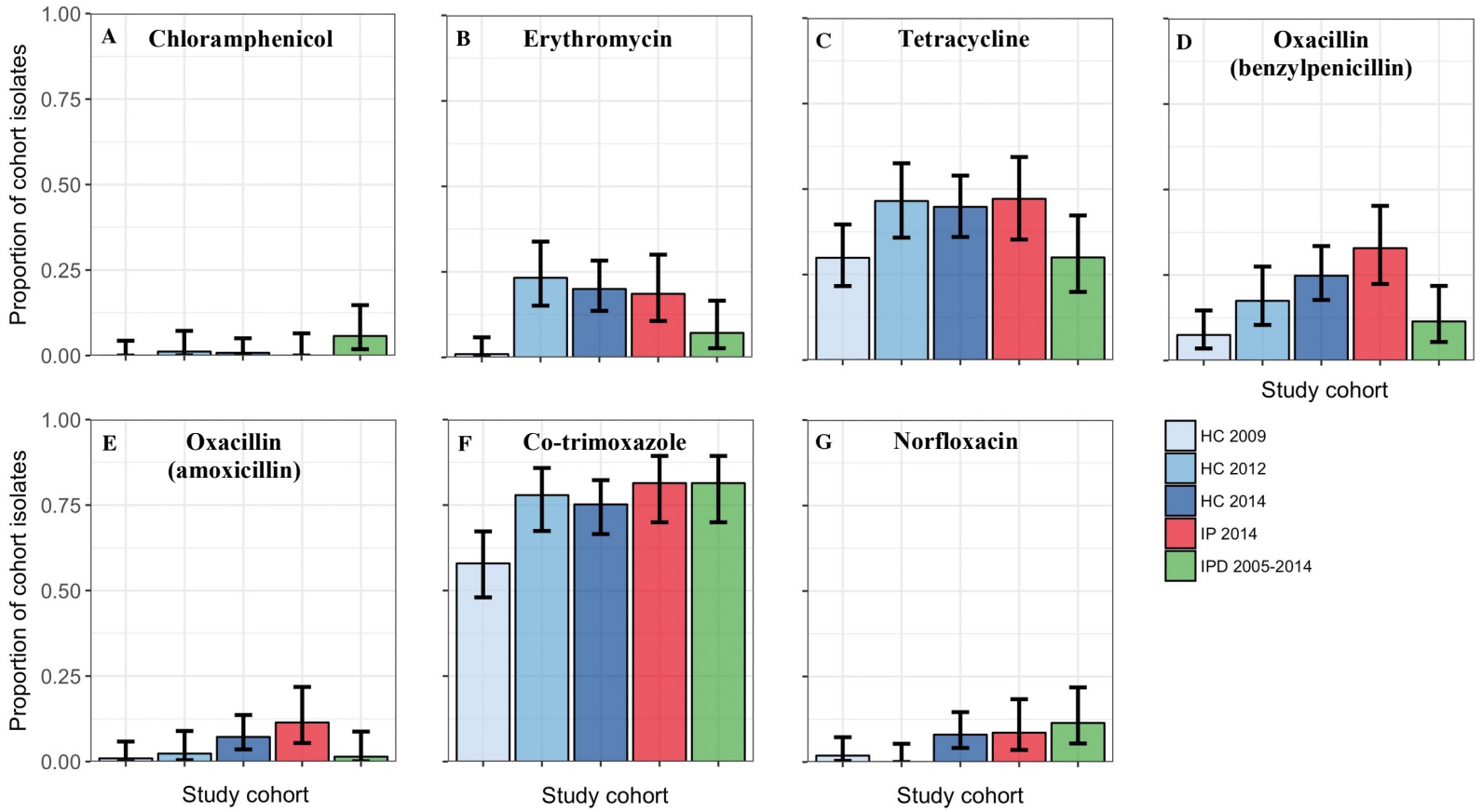


Figure 27. ***Proportion of pneumococcal samples amongst each cohort of Nepalese children that were non-susceptible to antimicrobials as determined by disc diffusion and EUCAST cut-points.***

Chloramphenicol (A), erythromycin (B), tetracycline (C), oxacillin (benzylpenicillin) (D), oxacillin (amoxicillin) (E), co-trimoxazole (F), and norfloxacin (G). HC 2009 (N=106), HC 2012 (N=84), HC 2014 (N=124), IP 2014 (N=69), and IPD 2005-2014 (N=68).

6.3.1 Pneumococcal non-susceptibility to chloramphenicol

Only 6/450 (1.3%) isolates were non-susceptible to chloramphenicol (figure 27A). Four of these were from the IPD cohort, three of which were serotype 1. A non-susceptible isolate was more likely to be identified in IPD isolates than healthy carriers (4/70 vs 2/314, p=0.0114).

catA encodes a chloramphenicol acetyltransferase that acts to modify chloramphenicol preventing it from binding to bacterial ribosomes.[194] There was a significant association with the presence of a *catA* gene and non-susceptibility to chloramphenicol (4/6 vs 2/452, p<0.0001).

6.3.2 Pneumococcal non-susceptibility to erythromycin

64/450 (14.2%) isolates were non-susceptible to erythromycin. Non-susceptibility was lowest in the HC 2009 cohort and significantly increased when compared with the HC 2012 and HC 2014 cohorts (1/106 vs 20/84, p<0.0001, and 1/106 vs 25/124 p<0.0001) (figure 27B).

ermB encodes a ribosomal methylase which acts to protect the binding site of the 23S portion of the ribosome from erythromycin binding. *mefA* and *msrD* encode efflux pumps which specifically act to remove macrolide molecules from the cell.[187] All but one isolate that was non-susceptible to erythromycin possessed a gene that conferred macrolide resistance. This relationship was statistically significant when compared with the susceptible isolates (63/64 erythromycin resistance genes detected in non-susceptible isolates versus 11/386 resistance genes detected in susceptible isolates, p<0.0001).

6.3.3 Pneumococcal non-susceptibility to tetracycline

182/450 (40.4%) isolates were non-susceptible to tetracycline. Non-susceptibility was lowest in the HC 2009 cohort and significantly increased subsequently in comparison with the HC 2012 and HC 2014 cohorts (32/106 vs 40/84, p=0.0163, and 32/106 vs 56/124 p=0.0213) (figure 27C).

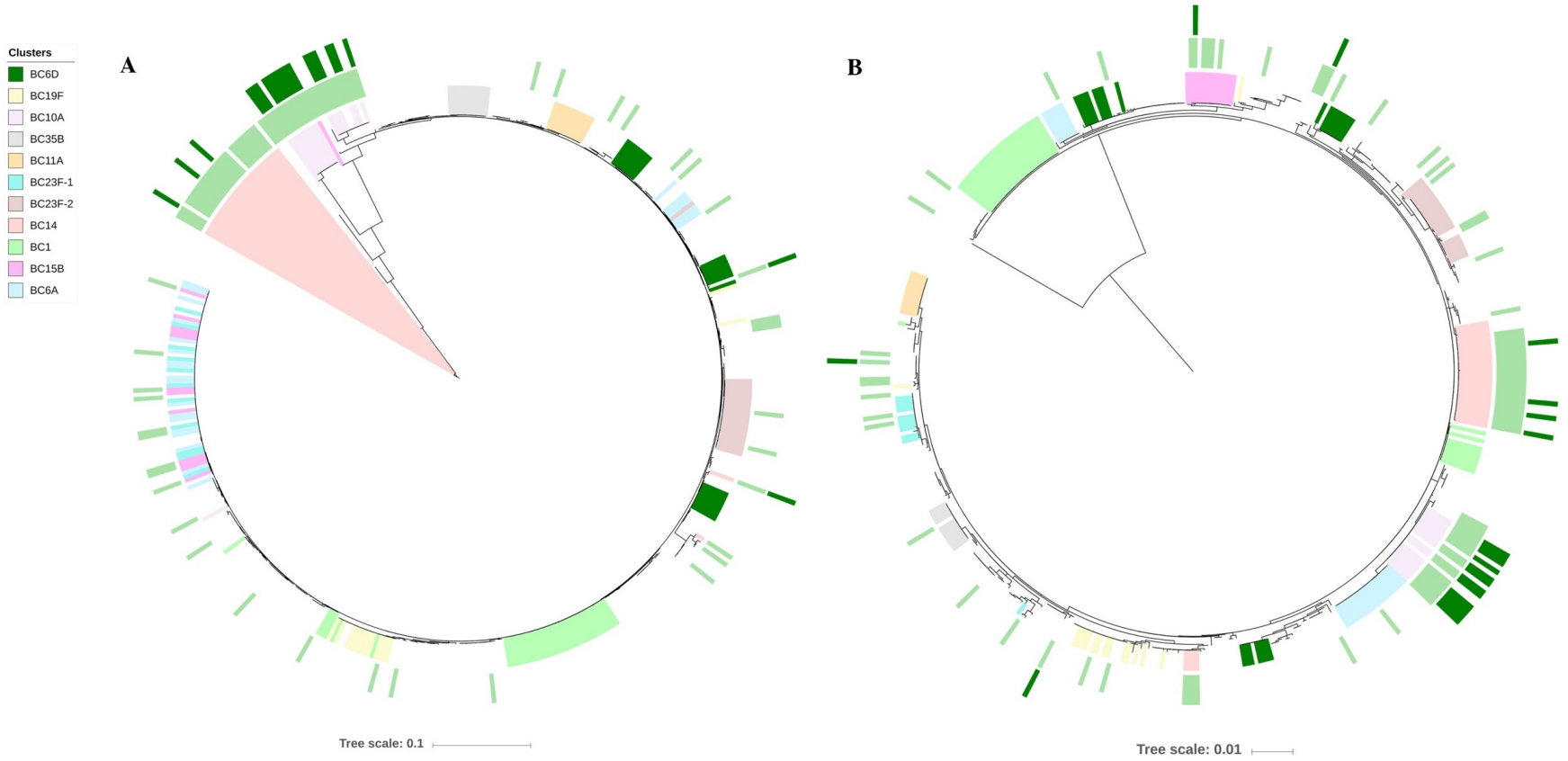
tet32, *tetO*, and *tetM* confer resistance to tetracycline by acting to dislodge tetracycline from its ribosomal binding site.[195] The majority of isolates that were non-susceptible to tetracycline had a gene detected which conferred tetracycline resistance. This relationship was statistically significant when compared to isolates that were susceptible (147/182 tetracycline resistance genes detected in non-susceptible isolates versus 11/268 tetracycline resistance genes found in susceptible isolates, $p<0.0001$).

6.3.4 Pneumococcal non-susceptibility to oxacillin

When applying the benzylpenicillin breakpoint to oxacillin disc testing, 85/450 (18.9%) of isolates were non-susceptible. A significant increase in non-susceptibility was noted when comparing the HC 2009 cohort to the HC 2012 (8/106 vs 15/84, $p=0.0427$) and HC 2014 cohorts (8/106 vs 31/124, $p<0.0004$) (figure 27D). Although there was an increase in non-susceptibility from the HC 2012 to HC 20140 cohort it was not a statistically significant change.

When applying the amoxicillin breakpoint to oxacillin disc testing, 21/450 (4.7%) of isolates were non-susceptible. A significant increase in non-susceptibility was noted when comparing HC 2009 with the HC 2014 cohort (1/106 vs 9/124, $p=0.0225$) (figure 27E).

Mosaicism of the penicillin binding proteins pbp1a, pbp2a, and pbp2x has been shown to confer non-susceptibility to beta-lactams. Examination of the phylogenies for *pbp1a*, *pbp2a*, and *pbp2x* in the present study showed that long branches related to beta-lactam non-susceptibility (figure 28).


Clusters
BC6D
BC19F
BC10A
BC35B
BC11A
BC23F-1
BC23F-2
BC14
BC1
BC15B
BC6A
A
Tree scale: 0.1
B
Tree scale: 0.01

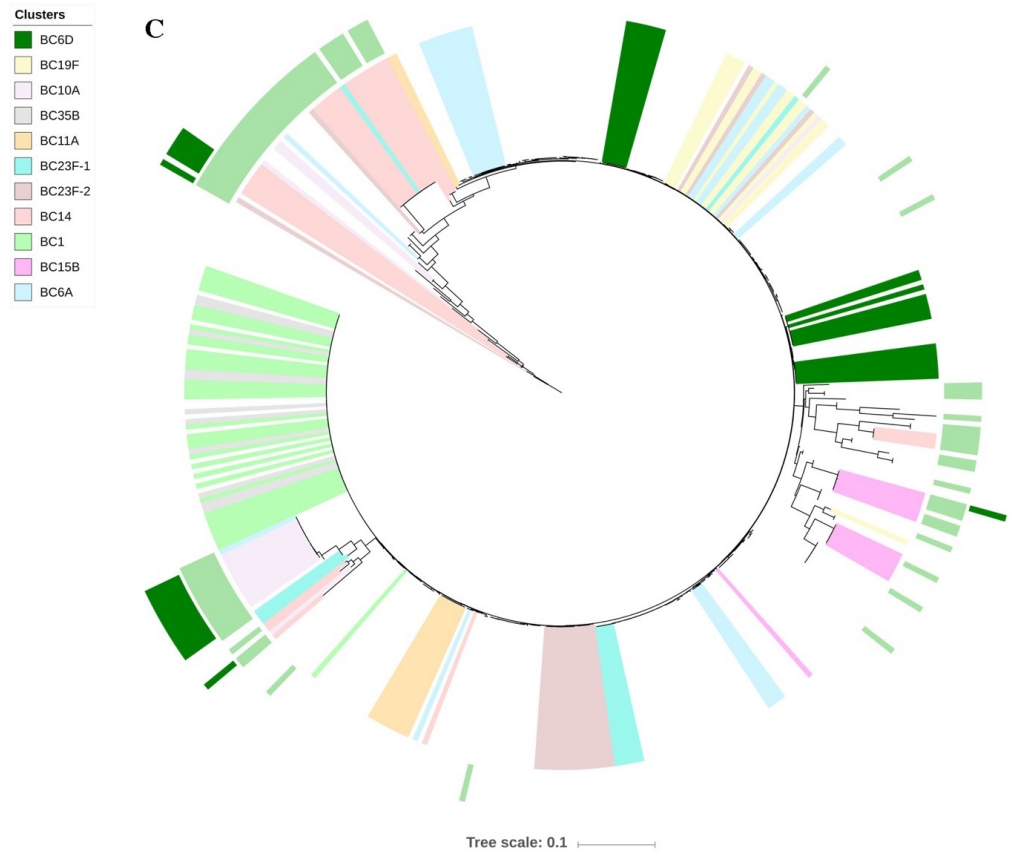


Figure 28. ***Maximum likelihood phylogeny of pbp1a (A), pbp2a (B), and pbp2x (C) amongst isolates collected from Nepalese children prior to PCV introduction, demonstrating the relationship of penicillin binding protein mosaicism to oxacillin non-susceptibility.***
The taxa belonging to the eleven largest primary clusters are highlighted. The light green coloured bars on the inner circle indicate those isolates which were found to be non-susceptible to oxacillin using the breakpoint which equated to benzylpenicillin non-susceptibility. The dark green coloured bars on the outer circle indicate those isolates which were found to be non-susceptible to oxacillin using the breakpoint which equated to amoxicillin non-susceptibility. The trees are rooted on the longest branch.

6.3.5 Pneumococcal non-susceptibility to co-trimoxazole

337/450 (74.9%) isolates were non-susceptible to co-trimoxazole. There was a significant increase in non-susceptibility from the HC 2009 to the HC 2012 cohort (62/106 vs 67/84, p=0.0018) (figure 27F). The two constituents of co-trimoxazole, Sulfamethoxazole and trimethoprim, each respectively act to inhibit the enzymes dihydropterate synthase (*folP*) and dihydrofolate reductase (*folA*). As such sequenced reads were mapped to the *Streptococcus pneumoniae* ATCC 700669 *folA* and *folP* genes. This showed that for both *folA* and *folP* phylogenies the longer branches related to non-susceptibility to co-trimoxazole (figure 29). Interestingly there is significant admixture of the detected alleles across the main population clusters.

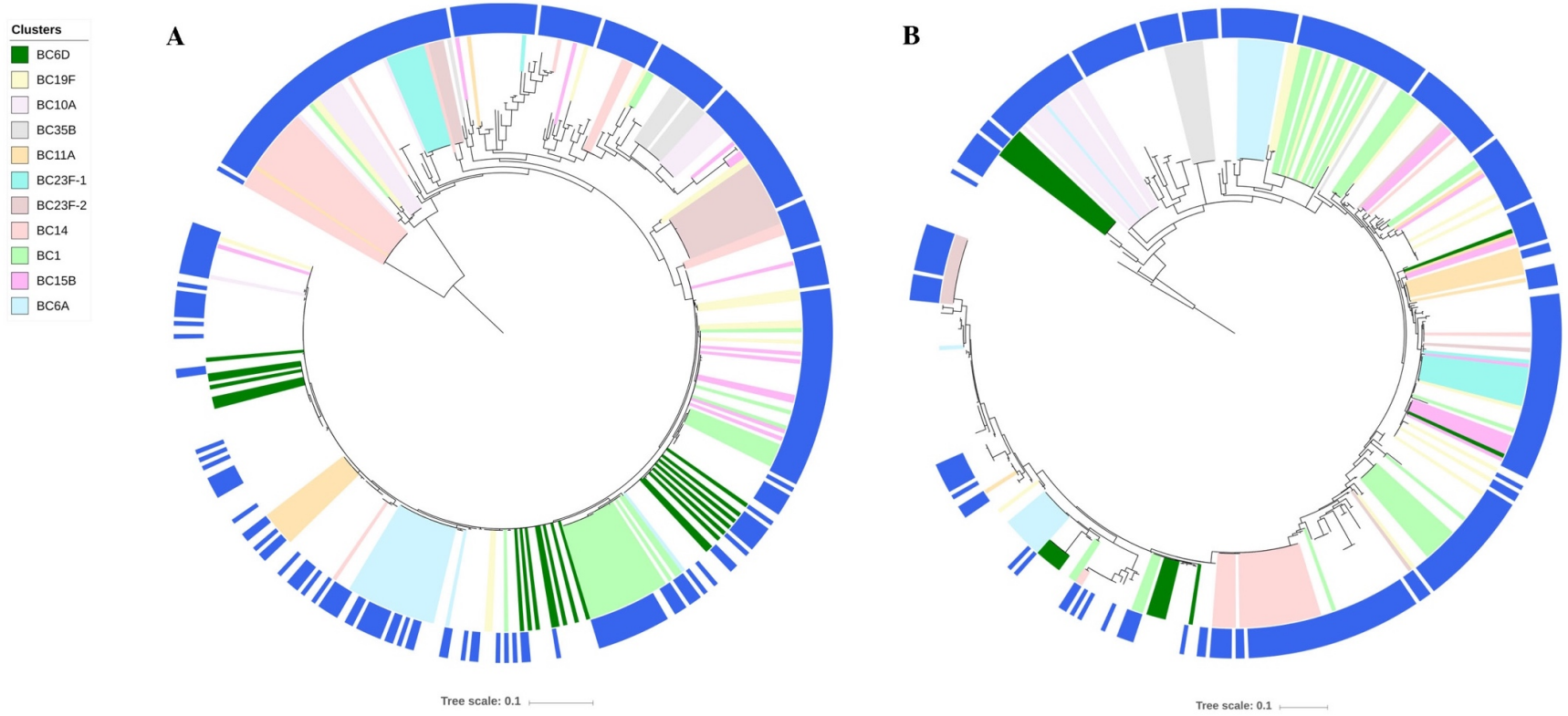


Figure 29. ***Maximum likelihood phylogeny of folA (A) and folP (B) amongst isolates collected from Nepalese children prior to PCV introduction.***
The taxa belonging to the eleven largest primary clusters are highlighted. The blue coloured bars on the outer circle indicate those isolates which were found to be non-susceptible to co-trimoxazole. The trees are rooted on the longest branch.

6.3.6 Pneumococcal non-susceptibility to norfloxacin

26/450 (5.8%) isolates were non-susceptible to norfloxacin. The prevalence of norfloxacin non-susceptibility was low in both the HC 2009 (2/106) and HC 2012 (0/84) cohorts and was significantly higher in the HC 2014 (10/124) cohort when compared with each of these earlier cohorts (HC 2009 vs HC 2014 p=0.0405, and HC 2012 vs HC 2014 p=0.0063) (figure 27G). The prevalence of norfloxacin non-susceptibility was also significantly higher in the IPD cohort compared with the combined HC 2009-2014 cohort (12/314 vs 8/68, p=0.0141).

As modification of the DNA gyrase mechanism, particularly g*yrA* and *parC*, have been described as a key mechanism for fluoroquinolone resistance in pneumococcus, the SNPs within these genes and their relationship to non-susceptibility were examined. When examining the phylogeny of *gyrA* it was evident that there was a cluster of non-susceptibility within BC1 (figure 30A).

This related to five serotype 1 isolates which had a SNP (C to T) at nucleotide position 242 which conferred an amino acid switch from serine to phenylalanine at amino acid position 81. Across the entire population this SNP was significantly associated with the presence of non-susceptibility (5/26 versus 0/424, p<0.0001). When examining the phylogeny of *parC* it was similarly evident that there was a group of BC1 isolates that were non-susceptible (figure 30B). This correlated with a SNP (C to A or T) at nucleotide position 200, with further examination across the data set demonstrating this SNP to be present in other norfloxacin non-susceptible isolates. This SNP conferred an amino acid switch at position 67 from serine to either a tyrosine (N=9) or phenylalanine (N=6). Both the tyrosine and phenylalanine types were significantly associated with norfloxacin non-susceptibility (8/26 versus 1/424, p<0.0001, and 6/26 versus 0/424, p<0.0001).

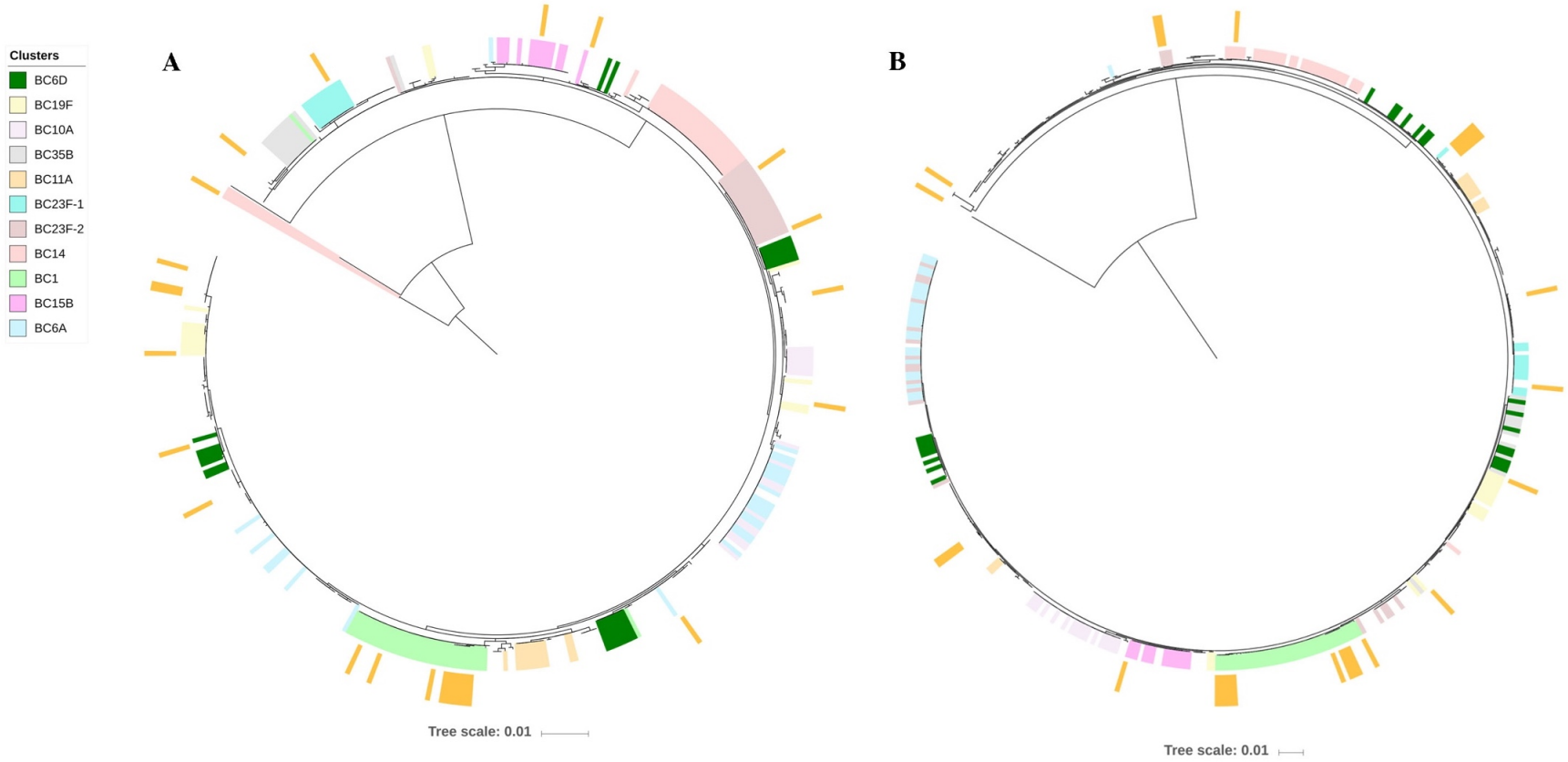


Figure 30. ***Maximum likelihood phylogeny of gyrA (A) and parC (B) amongst isolates collected from Nepalese children prior to PCV introduction.***
The taxa belonging to the eleven largest primary clusters are highlighted. The orange coloured bars on the outer circle indicate those isolates which were found to be non-susceptible to norfloxacin. The phylogenies are rooted on the longest branch.

6.4 Discussion

This is the largest study to examine pneumococcal antibiotic non-susceptibility in Nepal. This study demonstrates that pneumococcal antibiotic non-susceptibility is very prevalent, and is increasing, amongst Nepalese children in Kathmandu. Here it is shown that pneumococcal non-susceptibility to co-trimoxazole is so high that it is now clinically redundant to use this antibiotic for the treatment of pneumococcal disease. The rising trend in pneumococcal non-susceptibility to beta-lactams is also a concern as this class of antibiotic is now the first line treatment option in this setting. In general, the detection of previously described genes which confer non-susceptibility to antibiotics related closely with pneumococcal non-susceptibility in this setting.

6.4.1 Non-susceptibility to chloramphenicol is infrequent

Discussions with local paediatricians at Patan Hospital indicated that use of chloramphenicol was limited within the hospital and that there was very restricted access in the community. This is reflected in the low prevalence of non-susceptibility detected in this current study and is consistent with prior reports from Nepal (see table 33). *catA* was detected in this pneumococcal population and its presence correlated with non-susceptibility to chloramphenicol. As such if use of chloramphenicol were to increase it is likely that selection and subsequent proliferation of *catA* possessing chloramphenicol non-susceptible strains will occur.

6.4.2 Non-susceptibility to macrolides relates strongly to the detection of described macrolide resistance genes

Macrolides are the third most consumed antibiotic class in India and China, so it is likely that this is also the case in Nepal.[112] Particularly because macrolides have uses both in the treatment of enteric fever and pneumonia. An increase in the prevalence of non-susceptibility of pneumococci to erythromycin from 2009 to 2012 was seen in this current study and is

suggestive of significant exposure to macrolides in the community. The macrolide resistance genes *ermB*, *mefA*, and *msrD* typically reside within mobile genetic elements and thus are easily disseminated across bacterial populations.[187,196] In this study there is a very strong relationship between the presence of these three known macrolide resistance genes and non-susceptibility, thus making it unlikely there are other mechanisms conferring macrolide resistance currently in this population.

6.4.3 Pneumococcal non-susceptibility to tetracycline is high, but the clinical relevance is unclear

Although use of tetracyclines is reportedly less frequent than other classes of antibiotics in Nepal, this current study demonstrated high levels (182/450, 40.4%) of non-susceptibility amongst the samples tested. The presence of a previously described tetracycline resistance gene (*tet32*, *tetO*, and *tetM*) corresponded well with an isolate being phenotypically non-susceptible. However, there were 35/182 (19.2%) isolates that were non-susceptible, and did not have a detectable gene. This may be because there is an alternate mechanism for resistance that was not detected by the current studies means, for example SNPs or an undescribed resistance gene. Another possibility is that there were errors in assembly of the sequences leading to an under detection of the genes that were being screened.

6.4.4 Pneumococcal non-susceptibility to beta-lactams is increasing and needs to be monitored

The switch in policy in 2014, from using co-trimoxazole to amoxicillin as first line therapy for community acquired pneumonia amongst children, appears to correlate with the progressive increase over time in beta-lactam non-susceptibility in this present study. It has previously been shown that mosaicism of the penicillin binding proteins is related to beta-lactam non-susceptibility.[164] In this present study this also appears to be the case. Careful monitoring of

non-susceptibility to beta-lactams needs to continue in this setting as the high prevalence of non-susceptibility in isolates collected amongst children admitted with pneumonia is concerning.

6.4.5 Pneumococcal non-susceptibility to co-trimoxazole is very prevalent, making co-trimoxazole clinically redundant for the treatment of pneumococcal disease

The long-term and widespread use of co-trimoxazole has resulted in the development of extremely high levels of non-susceptibility in this population. These high levels of pneumococcal non-susceptibility to co-trimoxazole in this present study are consistent with other studies in the south Asia region.[164,197] Non-susceptibility to co-trimoxazole in this present study appears to be related to variation in *folA* and *folP*, consistent with another study in Maela. The fact that over three-quarters of IPD isolates are non-susceptible to co-trimoxazole underpins the need to avoid use of this antibiotic in this community for the treatment of any suspected pneumococcal disease.

6.4.6 Pneumococcal non-susceptibility to norfloxacin is highest amongst IPD isolates

Non-susceptibility to fluoroquinolones appears to be increasing, although currently remains at a relatively low prevalence compared with other antibiotic classes. Importantly the highest prevalence of non-susceptibility was amongst IPD isolates. Should the prevalence of non-susceptibility increase it may become clinically relevant. This is because fluoroquinolones are often used for the treatment of enteric fever in this setting, and in the context of antibiotics being dispensed by street-side pharmacies, pneumococcal disease may be mistreated as enteric fever if a limited medical history and examination is taken. In this study it was shown that the non-susceptibility to norfloxacin was related to two polymorphisms (one in *gyrA* and the other in *parC*) in the DNA gyrase mechanism.

6.4.7 Strengths and weaknesses of this study

The strengths of this study are the inclusion of isolates from a broad range of time and also the inclusion of carriage and disease sourced strains. Methodologically, assessment of non-susceptibility of all of the isolates was conducted in the same laboratory using the same guideline. Thus preventing any bias in reporting that would be present if assessment had been done progressively over the decade across which the isolates were collected. In addition, to date this study is the largest sample size of pneumococci from Nepal to have antibiotic resistance typing done.

Weaknesses of the study include; the resistance gene database used, and the approach to identifying molecular mechanisms of non-susceptibility. The database used is not specific to known pneumococcal resistance genes and is not completely exhaustive. It is conceivable that other resistance genes are present but were not detected because they were not in the database. However, in general there was a good relationship between the detection of a resistance gene and non-susceptibility. The approach used in this study, using short-read sequence typing (SRST), appears to have worked well for the identification of resistance genes. However not all antibiotic resistance is conferred by the presence of such genes, and other approaches such as genome-wide association studies may be a more practical approach to identifying polymorphisms that are associated with non-susceptibility.

Finally, care must be taken when comparing the patterns of antimicrobial non-susceptibility in the IPD cohort to the carriage cohorts, because the IPD isolates were collected over a very broad range of time compared with the carriage cohorts.

6.4.8 Conclusions

This study shows that pneumococcal non-susceptibility is increasing amongst Nepalese children. Of particular note is the fact that non-susceptibility to beta-lactams is increasing, and is particularly prevalent amongst carriage samples from children admitted with pneumonia.

This is concerning because beta-lactams are the first line treatment option for community acquired pneumonia, and growing levels of pneumococcal non-susceptibility may lead to increasing numbers of treatment failures.

This study also shows that previously described molecular mechanisms that confer non-susceptibility are present in this pneumococcal population. It is also shown that in some cases, previously undescribed mechanisms are present. For example, previously unreported polymorphisms in the DNA gyrase mechanism that are associated with fluoroquinolone non-susceptibility.

Ongoing monitoring of pneumococcal antibiotic non-susceptibility needs to be conducted in this present setting as there are likely to be clinically relevant implications. In addition, further steps towards regulation of the access to antibiotics needs to be considered.

Chapter 7 The global population structure of pneumococcus prior to PCV introduction

7.1 Introduction

It was soon apparent through carriage studies, that following the introduction of PCVs non-vaccine type pneumococci were replacing vaccine types.[198] In addition, it is also now being observed that disease incidence due to the expansion of NVTs is increasing and may soon result in overall pneumococcal disease incidence returning to pre-PCV levels.[90]

To understand how pneumococcus is able to adapt to PCV introduction it is fundamental to characterise the pneumococcal population prior to PCV introduction. This has been done in many regions by conventional serological typing and molecularly by MLST. One systematic review has assessed those studies reporting IPD serotypes from around the world.[199] This paper showed that serotypes 14, 6B, and 1, were the most commonly detected IPD serotypes globally. However, such studies consistently demonstrate geographical differences in pneumococcal serotype-specific distribution.[199] This is likely to be attributed to localised novel strain emergence, differing patterns of strain transmission between regions, and differing inter-strain patterns of competition.

It is important to note that serotype distribution does not necessarily reflect the underlying distribution of pneumococcal strains at a molecular level. As such many studies over the last decade have used MLST to characterise pneumococcal strains. MLST of pneumococcus has been used to characterise isolates based on the sequences of seven housekeeping genes.[200] Such studies have shown that different lineages, based on sequence type, can express different serotypes, that is these strains have undergone capsular switching.[201,202] Thus those strains of pneumococcus which possess a variety of capsules are capable of escaping immune selection due to PCV use.

Characterisation of pneumococcal strains at a global level by MLST is limited. There is no published meta-analysis which draws together regional sequence type distribution to give a global perspective. However, there is one study which has identified the sequence type of 426 isolates sourced over an 80-year period and across seven different countries. This study showed that recombination, not only of the capsular locus, is likely to have played an intrinsic role in the variation of strains.

Unfortunately, MLST is limited by the fact that it doesn't identify what the actual variations are that are responsible for adaptation within a strain. A key feature that needs to be identified if variation across pneumococcal strains is to be assessed. Whole genome sequencing can overcome these limitations however, with studies showing that this technology can be used on a large scale to characterise pneumococcal populations.[164]

7.1.1 The global pneumococcal sequencing project.

In order to characterise the adaptation of pneumococcus that occurs following PCV introduction the Global Pneumococcal Sequencing Project (GPS) was initiated, with funding from the Bill and Melinda Gates foundation and the Wellcome Trust. This project aims to conduct whole genome sequencing on approximately 20,000 pneumococcal isolates sourced from sites around the world before and after PCV introduction.[203] This project is thus the largest set of pneumococcal whole genome sequences collected to date and provides a remarkable opportunity to study the molecular characteristics of the global pneumococcal population prior to PCV introduction.

7.1.2 Rationale for a study

Therefore, using the GPS project database of pneumococcal whole genome sequences collected prior to PCV introduction it was aimed to describe the global population structure of pneumococcus prior to PCV introduction, to identify the proportion of each population cluster

that would be affected by a PCV, to identify strains which are unique to specific geographical regions, and to identify strains which have spread widely across the world.

7.2 Methods

7.2.1 Study design and objectives

Pneumococcal isolates that had undergone whole genome sequencing as part of the GPS project and were collected prior to PCV introduction in the country of collection were included in this study. The primary objective of this study is to determine and describe the global pneumococcal population structure prior to PCV introduction. Secondary objectives include; determining the diversity of strains which express serotypes contained in PCVs, determining which strains are unique to specific regions and those strains which are common across most regions, to characterise the specific strains which are unique or those which are especially common across most regions.

7.2.2 Sample selection, filtering, and QC

The Global Pneumococcal Sequencing project database accessed on the 19/12/2016. Samples which had passed the QC criteria described in section 5.2.1 were selected. These samples were then filtered according to the recorded vaccine period. Those samples which did not have a recorded vaccine period in the database but were collected in a year which was prior to nationwide introduction of PCV into the infant immunisation schedule, were classified as being from the pre-PCV period. Duplicate samples were also screened for and removed. Samples without a country of origin were excluded. Any country which contributed only had a single isolate was also removed from the data set.

7.2.3 *In silico* determination of serotype

The *in silico* serotype of each isolate was determined using a combination of two approaches. The first approach was to screen genome assemblies for capsular loci using an in house *in silico* PCR approach. The primers used were those previously published by the CDC.[204] The second approach was to map sequence reads to a reference database of capsular loci. Low mapping

percentages (under 50%) were considered indicative of capsular loss. Multiple high mapping percentages were classified as mixed samples. Multiple low matches to unrelated capsules were classified as mosaic. Serotype 6A and 6B isolates were classified as a 6E (6A) or 6E (6B) if there were found to possess the polymorphism in the *cps* locus consistent with the 6E genotype.[205]

7.2.4 Determination of sequence type

Sequence types (STs) were determined using MLST check which was used to compare the assembled genomes against the MLST database for *Streptococcus pneumoniae* (pubmlst.org/spneumoniae/).[206,207] All of the software developed by Pathogen Informatics at the Wellcome Trust Sanger Institute is freely available for download from GitHub under an open source license, GNU GPL 3.[206] Any novel profiles were characterised by short-read sequence typing and the novel profiles submitted to the MLST database (pubmlst.org/spneumoniae/).[208] goeBURST was applied to the set of STs in the dataset and clusters of STs descended from a recent common ancestor determined. These relationships were visualised using PHYLOviZ.[209,210]

7.2.5 Coarse mapping, generation of SNP alignment and hierBAPS clustering

As described in section 5.2.2, sequence reads from the selected isolates were mapped to the genome of *Streptococcus pneumoniae* ATCC 700669 using SMALT and then an alignment of SNP sites created. Hierarchical clustering was conducted on the alignment of SNPs as described in section 5.2.4 (computational time was approximately 56 days).

7.2.6 Phylogeny creation using FastTree.

An approximately-maximum-likelihood phylogenetic tree was generated from the SNP alignment using FastTree using a generalised time-reversible (GTR) model of DNA evolution.[211] The phylogeny and associated metadata was visualised using itol.embl.de.

7.2.7 Estimating recombination and mutation rates

As described in section 5.2.5 population clusters underwent fine mapping either with a previously described reference or a newly created custom reference genome (table 35) before recombination blocks and SNPs were determined.

Table 35. ***Details of references used for fine mapping of selected population clusters.***

Cluster name	Reference used	Serotype	Sequence type	Length, bp	Mean mapping percentage (range)	Accession number
BC6	Streptococcus_pneumoniae_23AF_ST172_v1.fa	23A/F	172	1897019	93.7 (91.1-96)	(in house WTSI)
BC8*	custom reference (lane 21127_1#89)	6B	12493	2064727	95.4 (86.2-99.8)	-
BC20	custom reference (lane 21127_2#82)	6E(6B)	12841	2090345	94.1 (91.3-99.2)	-
BC25	custom reference (lane 13415_4#17)	19A	2062	2122680	96.5 (94.3-99.4)	-
BC27	custom reference (lane 17794_7#100)	34	7319	2116535	97.2 (93.1-99.8)	-
BC33	custom reference (lane 12837_1#67)	6A	1094	2004987	99 (98.1-99.7)	-
BC34	custom reference (lane 15873_5#81)	6E(6B)	5516	2077039	98.7 (95.9-99.8)	-
BC37	custom reference (lane 15873_4#79)	23B	5706	2062697	99.1 (98.2-99.6)	-

*One isolate had a mapping percentage of 86.2%. The next lowest mapping percentage for an isolate was 92.9%.

7.2.8 Statistical analysis

CARTO was used to visualise the geographical distribution of the samples included in the analysis.[212] Simpson's index of serotype diversity for each population cluster was calculated using the formula:

$$\text{Simpson's index} = [\Sigma n(n-1)]/[N(N-1)]$$

Where: n = the number of times each serotype was identified. N = the total number of isolates within the population cluster. This returns a value between 0 and 1, with higher values indicating greater diversity.

Unless otherwise stated above all other visualisations and statistical analyses were performed using R version 3.3.1.[134]

7.3 Results

7.3.1 Overview of samples included in the analysis.

Following QC and filtering 7596 samples were included in the analysis. The majority of these samples (4020, 52.9%) were from the African continent, whilst blood was the single most common source of isolation of pneumococci (3279, 43.2%) (table 36). South Africa was the single country which contributed the largest number of samples (1301, 17.1%) (figure 30). PCVs were introduced into different settings at different times with samples included in this analysis being those that were collected before nation-wide PCV programmes were rolled out in each respective country. As such the time-period over which the samples were collected ranges from 1991 to 2016 (figure 31).

Table 36. ***Continent and host source of origin for pneumococcal isolates included in the analysis.***

Continent	NP swab	Blood	CSF	Invasive other	Other	Unknown
Africa	1433	1699	680	187	5	16
America	352	843	182	34	11	5
Middle East	1	23	43	5	35	0
Europe	35	190	165	4	97	0
Asia	691	503	62	171	7	93
Oceania	0	21	3	0	0	0

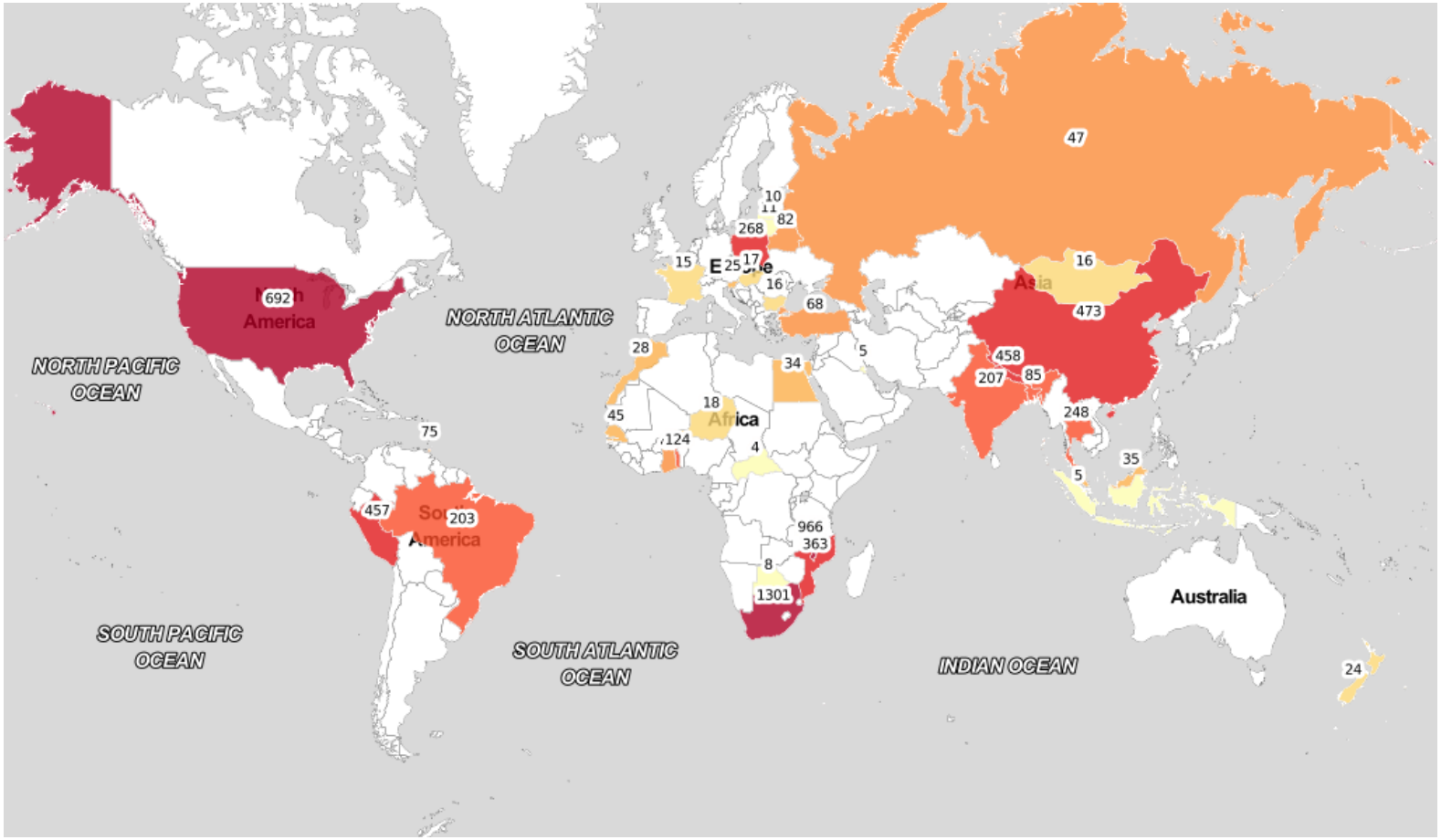


Figure 31. ***Geographical distribution of pneumococcal isolates included in the analysis.***
A total of 7596 pneumococcal samples collected prior to nation-wide PCV introduction underwent whole genome sequencing.

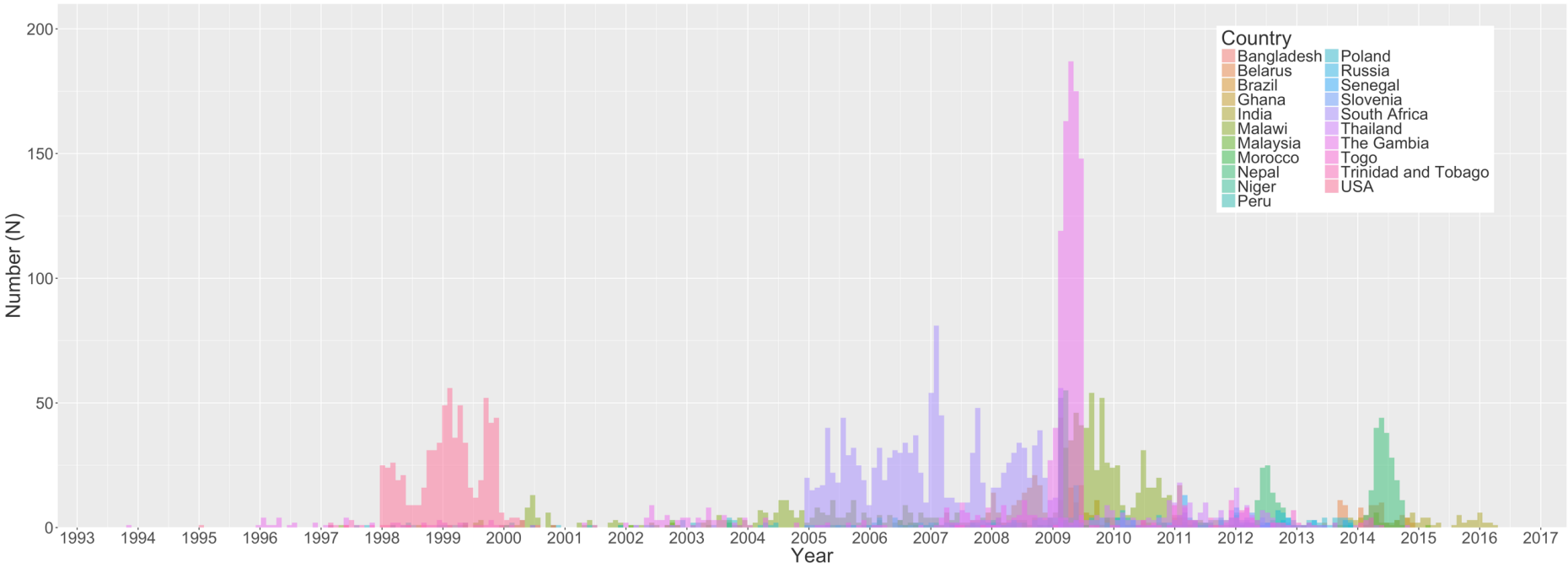


Figure 32. ***Temporal and regional distribution of pre-PCV era pneumococcal isolates selected for analysis.***
Of the 7596 samples which underwent whole-genome sequencing, 1741 did not have a complete date of collection and are not shown on this figure

7.3.1.1 Serotype and sequence type distribution of pneumococci prior to PCV introduction.

Serotypes 14 (814, 10.7%), 23F (576, 7.6%), and 19F (556, 7.3%) were the most common serotypes included in this dataset (figure 33). PCV7, PCV10, and PCV13 serotypes represented 2782 (36.6%), 3659 (48.2%), 4779 (62.9%), isolates respectively.

1910 distinct sequence types were identified. Sequence types 217, 289, and 156 were most frequently detected in this dataset and represented (239, 3.1%), (182, 2.4%), and (176, 2.3%) isolates respectively. ST217 consisted entirely of serotype 1 isolates, ST289 consisted entirely of serotype 5 isolates, whilst ST156 consisted of 119 serotype 14, 53 serotype 9V, two serotype 15B, one serotype 19F, and one serotype 23F isolates. A total of 871 clonal complexes were determined, 608/871 (69.8%) of these were singleton clonal complexes. ST217 and ST289 each resided within a singleton clonal complex which was unrelated to any other strains. ST156 was the most frequently identified ST that resided within a non-singleton clonal complex. ST156 resided at the centre of a large clonal complex (clonal complex 4) with numerous related strains (figure 34).

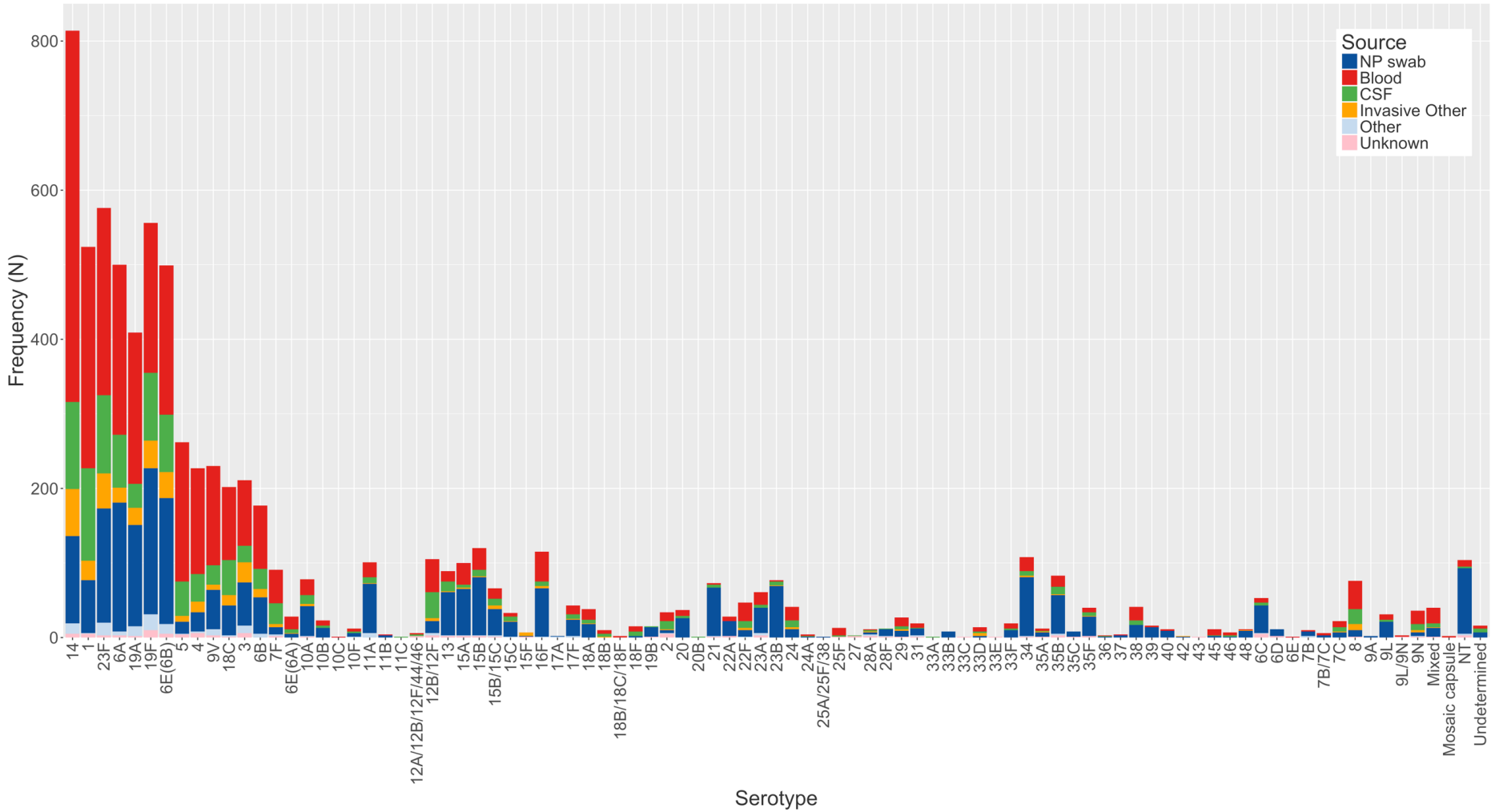


Figure 33. ***Serotype distribution and source of isolation of pneumococcal isolates included in the analysis of the global pneumococcal population prior to PCV introduction.*** 7596 pneumococcal isolates collected from across 36 different countries underwent whole-genome sequencing. The serotype of each isolate was determined by *in silico* analysis of the capsular locus. Each bar is coloured according to the number of samples that were isolated from a nasopharyngeal swab (blue), blood (red), CSF (green), other sterile fluid samples (orange), other sources (light blue), or un-reported (pink).

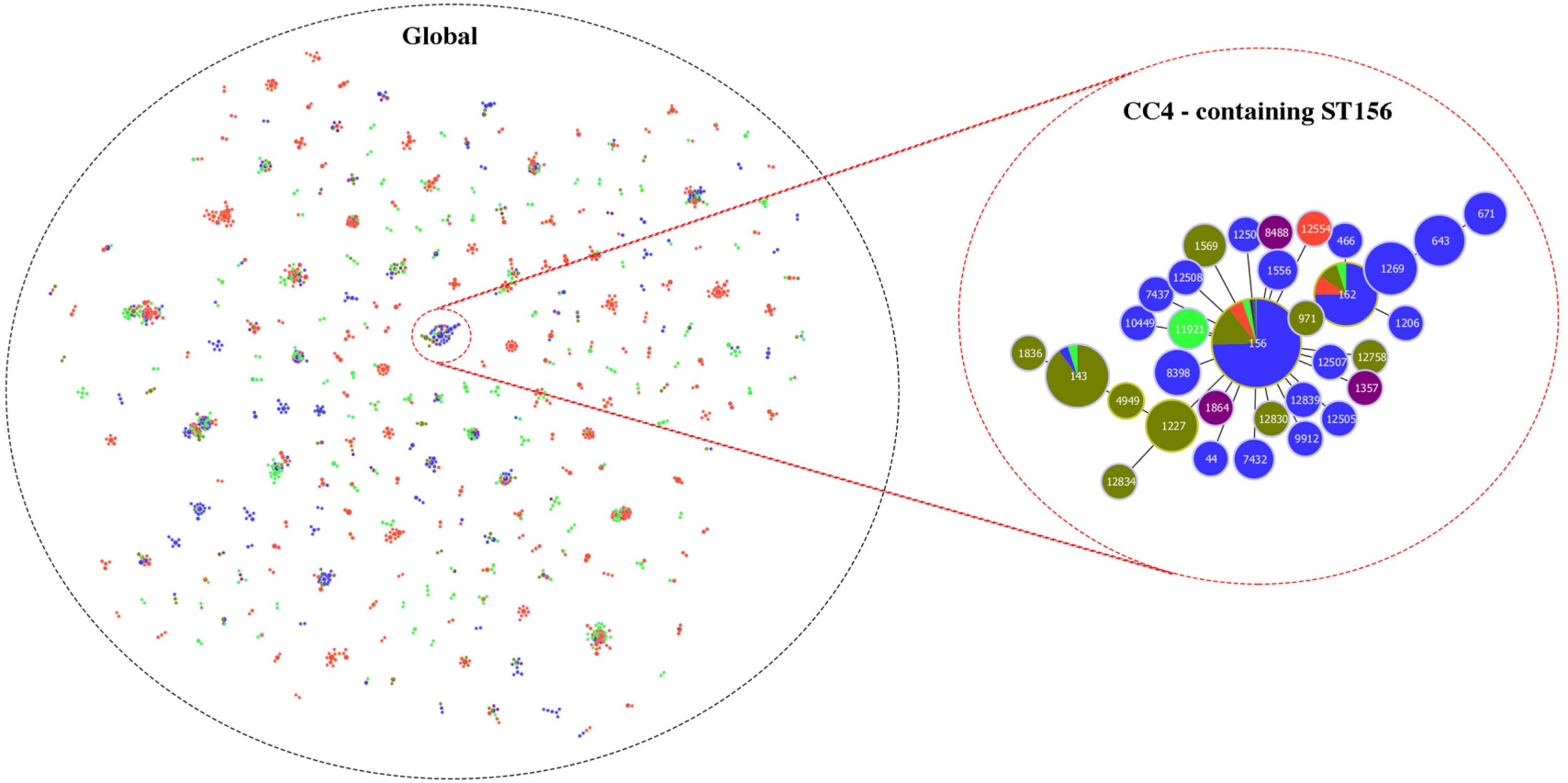


Figure 34. ***The global pneumococcal pattern of sequence type distribution prior to PCV introduction with a focus on clonal complex 4.***
Of the 1910 sequence types a total of 871 clonal complexes were determined. 608 clonal complexes were singletons. Clonal complex 4 contained sequence type 156 which is the most frequently represented sequence type that is not within a singleton clonal complex. Each sequence type circle is coloured according to the proportion of that sequence type represented by each continent (red = Africa, green = Asia, blue = America, olive = Europe, purple = Middle East, and cyan = Oceania).

7.3.2 Hierarchical BAPS clustering

Visualisation of the phylogeny generated based on the SNPs detected across the whole-genome of each isolate demonstrates that there are isolates which appear to cluster together based on geographic location (figure 35). Using a hierarchical clustering approach, a total of 45 primary population clusters were identified. The two largest clusters, clusters 3 and 4 were polyphyletic, and each represented 1155/7596 (15.2%) and 502/7596 (6.6%) isolates respectively (table 37). The largest monophyletic cluster was cluster 16 (492/7596, 6.5%) which consisted entirely of serotype 1 isolates. The primary clusters were then subsequently sub-categorised into a total of 217 secondary clusters and 335 tertiary clusters.

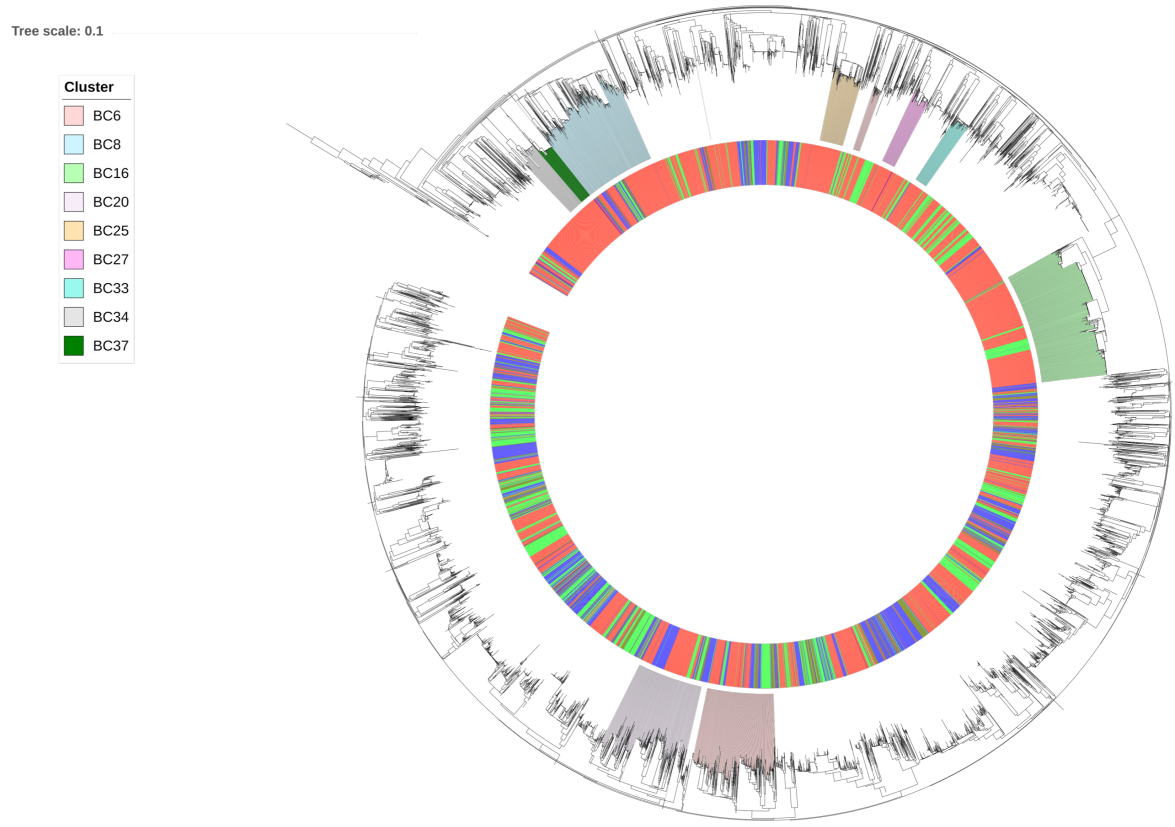


Figure 35. ***Whole-genome SNP informed maximum likelihood phylogenetic tree of pneumococcal isolates collected from across the world prior to PCV introduction.***
BC16 was the largest monophyletic population cluster. BC6, BC8, and BC20 were population clusters which were present in all continents. BC25, BC27, BC33, BC34, and BC37 were population clusters which were only present in a single continent. The inner circle is coloured according to the continent of origin (red = Africa, green = Asia, blue = America, olive = Europe, purple = Middle East, and cyan = Oceania). Note that this project is available online at: http://itol.embl.de/tree/192762721520111494712903

Table 37. ***Characteristics of each identified population cluster.***

Primary cluster	Number of isolates	Number of secondary clusters	Number of different STs	Number of different serotypes	Simpson's index of serotype diversity
1	276	3	34	6	0.50
2	99	7	24	2	0.42
3	1155	15	508	72	0.96
4	502	13	215	59	0.96
5	122	6	34	17	0.85
6	325	8	87	24	0.83
7	121	6	21	5	0.08
8	320	9	87	14	0.7
9	89	6	24	4	0.44
10	166	6	35	17	0.84
11	157	6	37	15	0.88
12	88	6	26	8	0.68
13	46	6	16	1	0
14	112	5	28	9	0.75
15	71	4	6	1	0
16	492	8	34	1	0
17	64	4	16	5	0.62
18	255	4	13	54	0.34
19	99	3	22	6	0.29
20	293	8	100	8	0.43
21	137	6	44	4	0.08
22	174	5	44	12	0.76
23	261	6	12	1	0
24	266	7	59	5	0.33
25	104	5	20	3	0.09
26	213	7	29	4	0.09
27	60	6	15	3	0.1
28	55	5	8	4	0.58
29	41	3	8	1	0
30	71	6	18	3	0.06
31	112	7	12	2	0.02
32	71	5	19	7	0.73
33	55	1	11	4	0.17
34	66	2	10	4	0.38
35	57	2	22	6	0.43
36	58	1	6	4	0.52
37	48	1	4	2	0.12
38	183	1	60	27	0.92
39	83	1	6	2	0.02
40	73	2	12	7	0.27
41	124	1	40	11	0.8
42	59	1	7	2	0.03
43	103	1	12	1	0
44	140	1	15	7	0.49
45	130	1	24	4	0.12

7.3.3 Strains which consist of a PCV serotype

The introduction of PCVs will result in the removal of lineages which express the capsule types covered by the vaccine from the population. However, some lineages contain more than one serotype and thus would have the ability to escape immune selection as a result of vaccine introduction. Within this dataset it was clear that some clusters such as clusters 16 and 23 which are represented entirely by serotype 1 and serotype 5 isolates respectively (thus have a Simpson's index of 0), clearly are limited in the ability to escape negative selection following vaccine introduction (figure 36). Other clusters however, such as clusters 6, 11 and 38 which are comprised of 65/324 (20.1%), 54/157 (34.4%), and 51/183 (27.9%) PCV13 isolates and have Simpson's indices of 0.83, 0.88, and 0.92 respectively, may be capable of escaping selection as a result of PCV introduction.

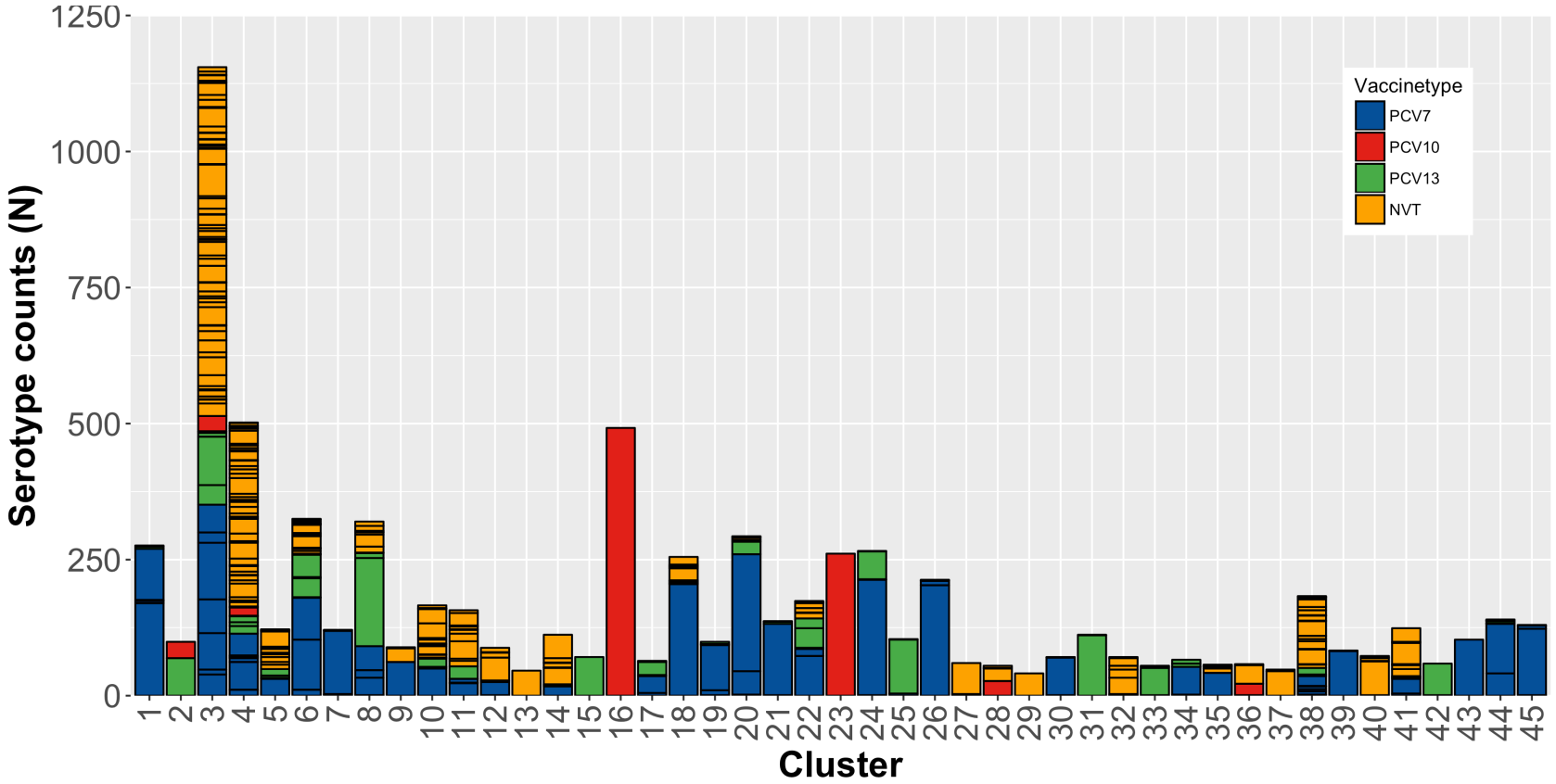


Figure 36. ***Proportion of each global pre-PCV pneumococcal population clusters covered by PCV7 (serotypes 4, 6B, 9V, 14, 18C, 19F, and 23F), the additional PCV10 (serotypes 1, 5, and 7F), and the additional PCV13 serotypes (serotypes 3, 6B, and 19A).***
7596 pneumococcal isolates underwent whole-genome sequencing, were mapped to the *Streptococcus pneumoniae* ATCC 700669 genome, and SNPs identified. SNPs across whole-genomes were then used to inform population clusters using the hierBAPS clustering tool.

7.3.4 Regional distribution of clusters

When examining the proportion of each cluster that was present in each country, it was apparent that some clusters were unique to only a single country or continent (figure 37). For example, cluster 34, a serotype 6E (6B) dominated lineage, was only detected in the Gambia. Similarly clusters 25 (serotype 19A dominated), 27 (serotype 34 dominated), 33 (serotype 6A dominated), and 37 (23B dominated) were only detected within the African continent. Interestingly cluster 35, a serotype 18C dominated lineage, was the only cluster not detected amongst those isolates collected from African countries.

It is important to note that the geographically restricted population clusters (BC25 N=104, BC27 N=60, BC33 N=55, BC34 N=66, and BC37 N=48) tend to be some of the smaller population clusters identified.

The monophyletic lineages which were distributed across the greatest number of countries were, clusters 6 (30 countries), 8 (23 countries), and 20 (22 countries). Clusters 6, 8 and 20 were serotype 6B, serotype 6A, and serotype 6E (6B) dominated lineages respectively.

Figure 37. ***Heatmap demonstrating the proportion of each pneumococcal population cluster that was sourced from each country.***
Countries are coloured according to continent (red = Africa, blue = America, green = Asia, purple = Europe, black = Middle East, and orange = Oceania).

7.3.5 Recombination and mutation events amongst selected population clusters that are geographically diverse or restricted

The regions of recombination and the SNPs which occur outside of these regions were determined for, the geographically diverse clusters BC6, BC8, and BC20, and for the geographically restricted clusters BC25, BC27, BC33, BC34, and BC37. For the geographically diverse populations clusters the mean r/m ratios ranged from 0.11 (95% CI 0.1-0.12) for BC6 to 0.14 (95% CI 0.12-0.16) for BC20 (figure 38). For the geographically restricted clusters the range of mean r/m ratios was much wider and spanned from 0.01 (95% CI 0-0.03) for BC37 to 0.2 (95% CI 0.17-0.25) for BC25. Interestingly BC34 and BC37 had the lowest mean r/m ratios and were also the least geographic diverse clusters, with BC34 only detected in The Gambia and all isolates except for one (which was found in Togo) from BC37 being detected also in The Gambia.

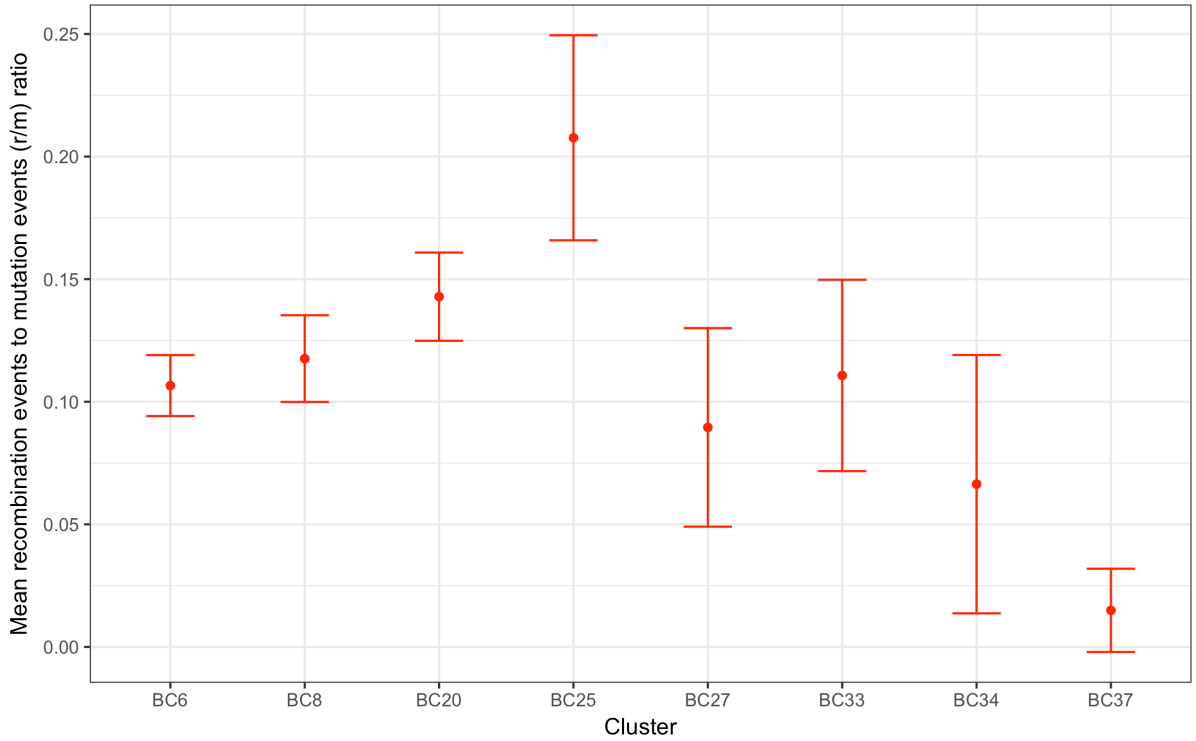


Figure 38. ***Mean recombination events to mutation events (r/m) in selected pneumococcal global population clusters and 95% CIs.***
Recombination and mutation events were determined for each isolate in each selected population cluster and the arithmetic mean calculated. BC6, BC8, and BC20 were geographically diverse populations. The remaining population clusters (BC25, BC27, BC33, BC34, and BC37) were geographically restricted.

7.4 Discussion

This is the first study to examine the global pneumococcal population using whole-genome sequencing. It shows that the majority of the pneumococcal population prior to PCV introduction can be classified into 43 monophyletic population clusters with the remaining isolates classified into one of two polyphyletic clusters. These population clusters are generally not limited by geographical borders with serogroup 6 pneumococci the most widely disseminated strains. Only five population clusters were identified within a single continent or country in this dataset. Interestingly many strains which contained isolates that express a PCV covered capsule also contained isolates that express NVT capsules. This would suggest that these strains would be capable of escaping selection as a result of PCV use.

7.4.1 Global population structure of pneumococci prior to PCV introduction

This study has classified the pre-PCV pneumococcal population into 45 clusters, where two of these clusters, representing 21.8% of all isolates, are polyphyletic. This is in comparison to a prior study of 3085 isolates from Maela, Thailand, which identified 33 clusters, with two of these, representing 17.2% of all isolates, polyphyletic clusters.[164] Clearly the increased number of clusters identified in the present study is probably due to the broader time-period, geography, and number of isolates analysed. Visualisation of the phylogeny suggests that one of the polyphyletic clusters (BC4) appears to have been defined on a set of common SNPs, but this set of SNPs is smaller than what the other monophyletic clusters are defined upon, hence the cluster spans across clades on the tree. The other polyphyletic cluster (BC3) is effectively undefined and is consistent with the 'bin' assignment that the clustering tool is programmed to allocate un-clustered isolates.

7.4.2 Serogroup 6 lineages were widely disseminated prior to PCV introduction

It is interesting to note that the most widely disseminated strains (BC6, BC8, and BC20) all belonged to serogroup 6. This may be because the serogroup 6 capsular polysaccharide has an intrinsic ability that facilitates spread. It may also be because these strains are resistant to antibiotics, as has been described in one study of serogroup 6 isolates from Spain.[213] Because BC6, BC8, and BC20 largely consist of serotypes which are covered by PCVs it is likely that the dominance of these strains across the globe will significantly diminish with the introduction of PCVs. It should also be noted that this would be further facilitated by the fact that immune responses to the polysaccharides of serotypes 6A and 6B cross-react with the serotype 6C pneumococci.[136]

7.4.3 Geographically restricted pneumococcal strains

Identification of geographically restricted strains is useful as the identification of any characteristics which limit spread may be useful for developing preventative interventions to pneumococcal disease. This study identified five lineages which were geographically restricted to a single continent. Notably all of these strains were limited to the African continent and all of these strains consisted of 104 or less isolates. This suggests that these strains may not have been detected in other settings simply because of a limitation in sampling or that they may be newly arisen from these settings. Further investigation of these strains over a longitudinal time-frame in the relevant setting may provide further insight into how these strains interact with other strains within these populations.

7.4.4 Strengths and weaknesses

The strengths of this study are; the unprecedentedly large sample size, and the balance of samples collected across different continents. This present study is almost double the size of the next largest study to assess a pneumococcal population at a whole-genome level. In

addition, this study draws upon samples sourced from numerous settings, and thus is truly representative of a global population.

Weaknesses of the study include; the time-period over which samples were collected, and the variation in environmental selective pressures. Because each country has introduced or is introducing PCVs at different times samples have been collected across different years prior to PCV introduction. This means that there may be changes in pneumococcal populations between countries between these different time-periods. In addition, each country has different levels of access to health care and as such influences such as antimicrobial use are variable across settings and may bias pneumococcal populations between countries. Finally, the host-genome across regions varies, and so variation in host immune selection may also be an unmeasured bias upon pneumococcal populations included in this study.

7.4.5 Conclusions

This study demonstrates that using whole-genome sequencing the global pneumococcal population prior to PCV introduction can be classified into 45 population clusters based on genetic sequence similarity. Importantly it is apparent that a number of cluster which contain PCV covered serotypes also contain NVTs and as such may be capable of progressively escaping the selective pressure of PCV introduction in the settings where these strains exist. Most strains were detected across geographical borders with only a selected few strains, of low frequency, restricted in geographic distribution. Further investigation of the characteristics that are associated with the 'spreadability' of pneumococcal strains is needed as the findings which arise may inform approaches to limiting the impact of those disease causing strains.

Chapter 8 Discussion

There are significant barriers to the ongoing prevention of pneumococcal disease around the world. This thesis details projects which are focused on addressing some of these key challenges. That is, in resource-rich settings such as the UK, which have PCVs well established within the infant immunisation schedule, challenges are faced by sub-optimal vaccine efficacy and serotype replacement. In resource-limited settings, such as Nepal, the key challenges are around accurately measuring local pneumococcal serotype distribution, quantifying the overall disease burden, and assessing vaccine effect in a setting without a national disease surveillance programme.

The progress in molecular technologies over the last decade mean that in resource-rich countries the opportunity to characterise pneumococci prior to PCV introduction have been largely missed. However, this is not the case in resource-limited settings where PCV programmes are only beginning to be instituted. As such isolates collected from these settings can be used to characterise the molecular epidemiology of these pneumococcal populations independent of the influence of PCVs on genetic selection. Thus providing the opportunity to gain key insights into pneumococcal genetic adaptability.

8.1 Carriage and IPD amongst UK children 5 years after PCV13 introduction

The introduction of PCV13 into the UK infant vaccination schedule has had a clear effect on the overall amount of asymptomatic carriage and invasive disease due to vaccine serotypes. However, the carriage, sero-prevalence, and IPD data presented in this thesis show that PCV13 effects at a serotype-specific level are differential, most notably for serotypes 3 and 19A. Further investigation into the underlying mechanisms of these differential effects, and whether intelligent use of PCV13, such as through alternate vaccine schedules, could mitigate these effects are needed. With regards to alternate vaccine schedules there are two main questions that need exploring.

Firstly, there is the question around the efficacy of a reduced dose, single dose prime and boost (1p+1), PCV13 schedule. As this schedule is currently being explored in a vaccine trial in the UK.[214] Prior studies on other PCV schedules indicate that the greater the amount of time between PCV doses the better the immunogenicity.[215] And so it may be that a 1p+1 schedule actually generates superior immunogenicity following boosting compared with the current two-dose prime and boost (2p+1) schedule in the UK.

Secondly there is the question around the utility of administering a further dose of PCV later in childhood. Such an approach would improve immunological persistence of antibody responses to the PCV13 polysaccharides and likely reduce the amount of vaccine type pneumococci circulating in this older age group secondary to waning immune responses to the infant PCV immunisation schedule.

Finally, the continued rise in incidence of NVT IPD amongst UK children and adults is a concerning trend and needs continued monitoring. Such monitoring, in the form of nation-wide surveillance and regional carriage studies, will allow the identification of those pneumococcal serotypes which are highly invasive and responsible for the bulk of disease. The capsular polysaccharides of these identified serotypes could then be included in a higher valence PCV.

8.2 Pneumococcal carriage in Nepal prior to PCV10 introduction and a comparison of conventional microbiological processing with DNA microarray for determining pneumococcal serotypes

This thesis demonstrates that pneumococcal carriage is frequent amongst healthy Nepalese children under 5 years of age in Kathmandu. Notably all pneumococcal serotypes covered by PCV10 were detected amongst this cohort prior to PCV10 introduction. Comparison of this cohort with a post-PCV cohort will provide a key insight into vaccine efficacy across this population, where IPD surveillance is limited.

The application of a DNA microarray technique to the detection of pneumococcal serotypes demonstrates that multiple-serotype carriage is frequent amongst Nepalese children carrying pneumococcus. Application of this technology in vaccine efficacy studies will provide an important insight into how the dynamics of multiple-serotype carriage are affected by PCVs. More importantly however, this technique provides a means for potentially reducing the sample size required for assessing vaccine efficacy across populations.

Although the microarray technology is specialised and not directly available in resource-limited settings, the initial microbiological processing and DNA extraction could be conducted at local sites and samples then shipped to a centralised specialist centre for microarray processing. Cost is also another important factor that may limit access to microarray technology, however it should be noted that conventional processing and Quellung serotyping have significant costs associated with staff time and anti-sera.

Overall the smaller samples sizes, reduced labour, and superior data output make the microarray platform the ideal option for the conduct of PCV efficacy studies.

8.3 Using *lytA* qPCR as a diagnostic tool for pneumococcal pneumonia in children

The importance of having an accurate diagnostic test for pneumococcal pneumonia cannot be underestimated. Should a good test become available it would allow more clinically focused treatment at the bedside, allow more concrete estimates of disease burden at an epidemiological level, and provide better insight into the assessment of PCV impact.

Assessment of the utility of *lytA* qPCR for diagnosing pneumococcal pneumonia in children showed that it had a good accuracy in differentiating children with pneumococcal pneumonia from children who did not meet the criteria for pneumococcal pneumonia but had an abnormal chest x-ray. Further refinement of the composite diagnostic used to define pneumococcal pneumonia in this thesis by the addition of additional data such as viral PCR results from nasopharyngeal swabs may further improve the diagnostic accuracy of *lytA* qPCR. Should this

prove to be the case then this assay would have a promising role in the clinical setting. This is further rationalised by the fact that the assay is a relatively straight forward PCR reaction which may be translated across to a user friendly and affordable platform that could be used in resource limited settings.

8.4 The molecular epidemiology of pneumococci amongst Nepalese children prior to PCV10 introduction

The assessment of the molecular epidemiology of pneumococci amongst Nepalese children prior to PCV10 introduction demonstrated that selective pressures on the pneumococcal genome were primarily due to host immunity and antibiotic use.

The selective pressures due to host immune responses are important to understand as they provide an insight into what the key virulence factors of pneumococci are. In this thesis it is shown that pneumococcal surface proteins, pspA and pspC, are under selective pressure. And that this appears to be a biologically plausible finding as both of these proteins are known to have functions which involve disabling components of the human innate immune system.[20]

It was also interesting to note that through the application of comparative genomics across three paired samples from blood and CSF it was possible to identify genes which are varied between the paired isolates environments. Although these data must be interpreted with caution because of the limited sample size. It is reassuring to note that those genes identified, such as the *ivr* locus and the pneumococcal histidine triad proteins, have biologically plausible functions in the transition from the blood to CSF setting.

In this thesis it was apparent that *pspA* and *pspC* have alleles which are predominantly found amongst disease isolates. In the case of the pneumococcal histidine triad proteins, *phtA* and *phtD*, it was also observable that there is a reasonable level of gene homology across the analysed population. As such specific pspA and pspC sub-types along with pneumococcal histidine triad proteins make plausible candidate antigens for a pneumococcal vaccine which

would not be limited by serotype. Animal studies of immunogenicity and animal challenge models of pneumococcal disease using these antigens would be the next stage of developing such a vaccine.

8.5 The pattern of pneumococcal antibiotic resistance amongst Nepalese children prior to PCV10 introduction

Assessment of pneumococcal antibiotic non-susceptibility amongst Nepalese children demonstrates two concerning trends. Firstly, pneumococcal non-susceptibility to co-trimoxazole is so high that the use of this antibiotic to treat suspected pneumococcal disease would be impractical. Secondly, pneumococcal non-susceptibility to beta-lactams is increasing in this population, in keeping with the change in clinical guidance which now recommends the use of beta-lactams as first line therapy for community acquired pneumonia.

Targeted screening and examination of the whole genome sequence data and comparison with the information on antibiotic susceptibility allowed the determination of which molecular mechanisms were responsible for antibiotic non-susceptibility. In the pneumococcal population examined in this thesis, it is apparent that previously described genes which confer antibiotic non-susceptibility are primarily responsible for pneumococcal non-susceptibility to macrolides, tetracyclines, and chloramphenicol. Target protein modification as a process to confer antibiotic non-susceptibility was also evident for isolates that were non-susceptible to beta-lactams, co-trimoxazole, and fluoroquinolones.

Looking forward, PCV10 introduction is likely to reduce the prevalence of antibiotic non-susceptibility in the short term, particularly for; beta-lactams where non-susceptibility is dominated by serotype 14 isolates, and fluoroquinolones where non-susceptibility is dominated by serotype 1 isolates. However, serotype replacement will occur eventually following PCV10 introduction and as such the patterns of antibiotic non-susceptibility in these emerging strains needs to be followed. In addition, continued assessment of pneumococcal antibiotic resistance

patterns in this setting is fundamental to informing clinical decisions on antibiotic treatment and should also be used to lobby for better regulation of antibiotic use.

8.6 The global population structure of pneumococci prior to PCV introduction

The analysis conducted in this thesis of the whole-genome sequences of the global pneumococcal population prior to PCV introduction is the largest such analysis to do done to date. It shows that although horizontal gene transfer amongst pneumococci creates difficulties in clustering isolates together based on sequence similarity, this challenge can be overcome and population clustering successfully performed on a large dataset. Albeit with the use of extreme levels of computing resources and time. The realisation that many population clusters that contain PCV serotypes also contain NVTs underpins the need to examine how these strains are affected in the post-PCV era.

This thesis shows that different strains of pneumococci fall along a spectrum of 'spreadability' with some strains appearing to be better at spreading across geographical borders than others. Further exploration of what the underlying genetic factors are which relate to the ability of some strains to spread better than others is needed as this may inform intelligent approaches to preventing disease strain transmission. Further analysis which may be done on this current data set to approach such a question would be to conduct a genome-wide association study looking to see if there is any association between specific genes and/or gene variations which are associated with transmissibility.

8.7 Further work

8.7.1 Understanding pneumococcal carriage amongst UK children 7 years after PCV13 introduction

IPD surveillance over the last year in the UK has shown an increase in serotype 19A disease, most notably amongst older adults. This observation coupled with the sub-optimal effect of

PCV13 on carriage and disease amongst UK children described in this thesis have led to the start-up of a further pneumococcal carriage study amongst UK children which has now commenced recruitment (ClinicalTrials.gov Identifier: NCT03102840). This study aims to recruit 1600 children aged 13-48 months who have received three doses of PCV13. With the primary objective being the prevalence of serotype 19A carriage amongst the recruited children. The sample size was selected based on the study having 80% power to detect an increase in serotype 19A carriage from 0.91% to 1.62% at a significance of $p<0.05$, when comparing this new study cohort with the one described in the second chapter of this thesis.

8.7.2 Collection of pneumococcal isolates amongst Nepalese children following the introduction of PCV10

Collection of pneumococcal isolates amongst Nepalese children in Patan, Kathmandu, Nepal, following introduction of PCV10 is currently underway. This includes the ongoing sampling and assessment of healthy carriage, carriage amongst children with pneumonia, and invasive pneumococcal disease. These samples will be analysed in the same way as the pre-PCV10 samples described in this thesis and thus will provide the perfect complement of data to draw detailed insights of vaccine efficacy.

These analyses will include; the assessment of the serotype-specific distribution of pneumococci amongst Nepalese children by conventional microbiology and microarray, the whole-genome sequencing of pneumococcal isolates collected in the post-PCV era, and ongoing assessment of pneumococcal antimicrobial resistance patterns.

8.7.3 Further development of real-time PCR of *lytA* from nasopharyngeal samples as a diagnostic test for pneumococcal pneumonia

The use of carriage density to inform the diagnosis of pneumococcal carriage described in this thesis shows promise as a diagnostic test that could be used in the field. In order to further

refine and develop this test it is important to conduct the following. Firstly, the sample size over which the assessment of carriage density has been conducted should be increased. This will allow a higher resolution assessment of the diagnostic capability of the test, as there will be more cases of pneumococcal disease within the cohort. Secondly the acquisition of further clinical data such as serum C-reactive protein levels may help further define pneumococcal/bacterial disease cases from non-pneumococcal/bacterial disease cases. Finally, further assessment of clinical samples, such as the nasopharyngeal swabs for the presence of other respiratory pathogens, for example influenza and respiratory syncytial virus would further aid in the definition of pneumococcal disease cases and provide greater certainty in the performance of the test.

8.7.4 Development of a viral vector based vaccine containing pneumococcal surface proteins

The surface proteins identified as being under host selective pressure in chapter 5 of this thesis have been used as the premise for a grant to fund the investigation of these proteins within a viral vector based vaccine.

The aim of this grant is to insert pneumococcal histidine triad proteins, pneumococcal surface protein C, and pneumococcal surface protein A, into an adenovirus vector. This viral construct would then be tested in mice for its ability to generate antibody responses to the included pneumococcal proteins. If this were to be successful, then the utility of this vaccine in preventing pneumococcal infection could then be subsequently tested in mice.

Should such a vaccine progress through the phases of vaccine development then it might be envisioned that its utility would be in conjunction with PCVs. This is because cessation of PCV programmes, which are increasingly established around the world, may result in a resurgence of PCV covered disease.

References

1. Sternberg, G. M. A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva. An experimental research. *Natl Bd Heal. Bull* **2,** 781–3 (1881).
2. Pasteur, L., Chambaland, C. & Roux, E. Sur une maladie nouvelle, provoquée pnr la salive d'un enfant mort de la rage. *C. R. Acad. Sci. Hebd. Seances Acad. Sci. D.* **xcii,** 159 (1881).
3. Gram, H. C. Ueber die isolierte Farbung der Schizomyceten in Schnitt-und Trockenpraparaten. *Fortschr. Med.* **2,** 185–189 (1884).
4. Netter, A. De l'endocardite vegetante-ulcereuse d'origine pneumonique. *Arch. Physiol. Norm. Pathol.* **3,** 106–161 (1886).
5. Netter, A. De la meningite due au pneumocoque (avec ou sans pneumonie). *Arch. Gen. Med.* **7,** 434–455 (1887).
6. Weichselbaum, A. Ueber seltenere Localisationen des pneumonischen Virus (Diplococcus pneumoniae). *Wien. Klin. Wochenschr.* **2,** 659–661 (1888).
7. Klemperer, G. & Klemperer, F. Versuche über Immunisierung und Heilung bei Pneumokokkeninfektion. *Berliner Klin. wochenschrift* **28,** 833–5, 869–75 (1891).
8. Neufeld, F. Ueber die Agglutination der Pneumokokken und über die Theorie der Agglutination. *Zeitschrift Fur Hyg. Und Infekt.* **40,** 54–72 (1902).
9. Neufeld, F. Ueber ein specifische bakteriolytische Wirkung der Galle. *Zeitschrift Fur Hyg. Und Infekt.* **34,** 454–464 (1900).
10. Neufeld, F. & Haendel, L. Weitere Untersuchungen über Pneumokokken Heilsera: III Mitteilung: Ueber Vorkommen und Bedeutung atypischer Varietäten des Pneumokokkus. *Arb Kais. Gesundheitsamte* **34,** 293–304 (1910).
11. McLeod, J. W. & Gordon, J. Production of Hydrogen Peroxide by Bacteria. *Biochem. J.* **16,** 499–506 (1922).
12. Pikis, A., Campos, J. M., Rodriguez, W. J. & Keith, J. M. Optochin resistance in Streptococcus pneumoniae: mechanism, significance, and clinical implications. *J. Infect. Dis.* **184,** 582–90 (2001).
13. Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. The Streptococcus pneumoniae Capsule Inhibits Complement Activity and Neutrophil Phagocytosis by Multiple Mechanisms. *Infect. Immun.* **78,** 704–715 (2010).
14. Park, I. H. *et al.* Discovery of a new capsular serotype (6C) within serogroup 6 of Streptococcus pneumoniae. *J. Clin. Microbiol.* **45,** 1225–33 (2007).
15. *Streptococcus pneumoniae: textbook in Serotyping, Virulence Factors and Enzyme-linked Immunosorbent Assay (ELISA) for Measuring Pneumococcal Antibodies*. (Statens Serum Institut, 2013).
16. Skov Sørensen, U. B., Blom, J., Birch-Andersen, A. & Henrichsen, J. Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infect. Immun.* **56,** 1890–6 (1988).
17. Sørensen, U. B. S., Henrichsen, J., Chen, H.-C. & Szu, S. C. Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of Streptococcus pneumoniae revealed by immunochemical methods. *Microb. Pathog.* **8,** 325–334 (1990).
18. Kamerling, J. in *Streptococcus pneumoniae Molecular Biology and Mechanisms of Disease* 81–114 (2000).
19. Bui, N. K. *et al.* Isolation and analysis of cell wall components from Streptococcus pneumoniae. *Anal. Biochem.* **421,** 657–666 (2012).
20. Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of Streptococcus

pneumoniae virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* **6,** 288–301 (2008).
21. Hammerschmidt, S., Bethe, G., Remane, P. H. & Chhatwal, G. S. Identification of pneumococcal surface protein A as a lactoferrin-binding protein of Streptococcus pneumoniae. *Infect. Immun.* **67,** 1683–7 (1999).
22. Tu, A. H., Fulgham, R. L., McCrory, M. A., Briles, D. E. & Szalai, A. J. Pneumococcal surface protein A inhibits complement activation by Streptococcus pneumoniae. *Infect. Immun.* **67,** 4720–4 (1999).
23. Hammerschmidt, S., Talay, S. R., Brandtzaeg, P. & Chhatwal, G. S. SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol. Microbiol.* **25,** 1113–24 (1997).
24. Cheng, Q., Finkel, D. & Hostetter, M. K. Novel purification scheme and functions for a C3-binding protein from Streptococcus pneumoniae. *Biochemistry* **39,** 5450–7 (2000).
25. Dave, S., Brooks-Walter, A., Pangburn, M. K. & McDaniel, L. S. PspC, a Pneumococcal Surface Protein, Binds Human Factor H. *Infect. Immun.* **69,** 3435–3437 (2001).
26. Iovino, F., Seinen, J., Henriques-Normark, B. & van Dijl, J. M. How Does Streptococcus pneumoniae Invade the Brain? *Trends Microbiol.* **24,** 307–315 (2016).
27. Mosser, J. L. & Tomasz, A. Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *J. Biol. Chem.* **245,** 287–98 (1970).
28. Gilbert, R. J. *et al.* Two structural transitions in membrane pore formation by pneumolysin, the pore-forming toxin of Streptococcus pneumoniae. *Cell* **97,** 647–55 (1999).
29. Berry, A. M., Lock, R. A. & Paton, J. C. Cloning and characterization of nanB, a second Streptococcus pneumoniae neuraminidase gene, and purification of the NanB enzyme from recombinant Escherichia coli. *J. Bacteriol.* **178,** 4854–60 (1996).
30. Austrian, R. Morphologic variation in pneumococcus. I. An analysis of the bases for morphologic variation in pneumococcus and description of a hitherto undefined morphologic variant. *J. Exp. Med.* **98,** 21–34 (1953).
31. Weiser, J. N., Austrian, R., Sreenivasan, P. K. & Masure, H. R. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect. Immun.* **62,** 2582–9 (1994).
32. Manso, A. S. *et al.* A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat. Commun.* **5,** 5055 (2014).
33. Techasaensiri, B., Techasaensiri, C., Mejías, A., McCracken, G. H. & Ramilo, O. Viral Coinfections in Children With Invasive Pneumococcal Disease. *Pediatr. Infect. Dis. J.* **29,** 1 (2009).
34. Smith, A. M. *et al.* Kinetics of Coinfection with Influenza A Virus and Streptococcus pneumoniae. *PLoS Pathog.* **9,** e1003238 (2013).
35. Winkelstein, J. A., Abramovitz, A. S. & Tomasz, A. Activation of C3 via the alternative complement pathway results in fixation of C3b to the pneumococcal cell wall. *J. Immunol.* **124,** 2502–6 (1980).
36. Mook-Kanamori, B. B., Geldhoff, M., van der Poll, T. & van de Beek, D. Pathogenesis and pathophysiology of pneumococcal meningitis. *Clin. Microbiol. Rev.* **24,** 557–91 (2011).
37. Taryle, D. A., Potts, D. E. & Sahn, S. A. The incidence and clinical correlates of parapneumonic effusions in pneumococcal pneumonia. *Chest* **74,** 170–3 (1978).
38. Strange, C. & Sahn, S. A. The definitions and epidemiology of pleural space infection.

Semin. Respir. Infect. **14,** 3–8 (1999).
39. Shivshankar, P., Boyd, A. R., Le Saux, C. J., Yeh, I.-T. & Orihuela, C. J. Cellular senescence increases expression of bacterial ligands in the lungs and is positively correlated with increased susceptibility to pneumococcal pneumonia. *Aging Cell* **10,** 798–806 (2011).
40. Robinson, K. A. *et al.* Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era. *JAMA* **285,** 1729–35 (2001).
41. Said, M. A. *et al.* Estimating the Burden of Pneumococcal Pneumonia among Adults: A Systematic Review and Meta-Analysis of Diagnostic Techniques. *PLoS One* **8,** e60273 (2013).
42. Korppi, M., Koskela, M., Jalonen, E. & Leinonen, M. Serologically indicated pneumococcal respiratory infection in children. *Scand. J. Infect. Dis.* **24,** 437–43 (1992).
43. Nohynek, H. *et al.* The causes of hospital-treated acute lower respiratory tract infection in children. *Am. J. Dis. Child.* **145,** 618–22 (1991).
44. Wolfe, W. G., Spock, A. & Bradford, W. D. Pleural fluid in infants and children. *Am. Rev. Respir. Dis.* **98,** 1027–32 (1968).
45. Chonmaitree, T. & Powell, K. R. Parapneumonic pleural effusion and empyema in children. Review of a 19-year experience, 1962-1980. *Clin. Pediatr. (Phila).* **22,** 414–9 (1983).
46. Freij, B. J., Kusmiesz, H., Nelson, J. D. & McCracken, G. H. Parapneumonic effusions and empyema in hospitalized children: a retrospective review of 227 cases. *Pediatr. Infect. Dis.* **3,** 578–91
47. Avni, T., Mansur, N., Leibovici, L. & Paul, M. PCR Using Blood for Diagnosis of Invasive Pneumococcal Disease: Systematic Review and Meta-Analysis. *J. Clin. Microbiol.* **48,** 489–496 (2010).
48. Lahti, E. *et al.* Pneumolysin polymerase chain reaction for diagnosis of pneumococcal pneumonia and empyema in children. *Eur. J. Clin. Microbiol. Infect. Dis.* **25,** 783–789 (2006).
49. Le Monnier, A. *et al.* Microbiological Diagnosis of Empyema in Children: Comparative Evaluations by Culture, Polymerase Chain Reaction, and Pneumococcal Antigen Detection in Pleural Fluids. *Clin. Infect. Dis.* **42,** 1135–1140 (2006).
50. Pride, M. W. *et al.* Validation of an immunodiagnostic assay for detection of 13 Streptococcus pneumoniae serotype-specific polysaccharides in human urine. *Clin. Vaccine Immunol.* **19,** 1131–41 (2012).
51. Bonten, M. J. M. *et al.* Polysaccharide Conjugate Vaccine against Pneumococcal Pneumonia in Adults. *N. Engl. J. Med.* **372,** 1114–1125 (2015).
52. Alere BinaxNOW Streptococcus pneumoniae Antigen Card - Alere. at <http://www.alere.com/en/home/product-details/binaxnow-streptococcus-pneumoniae.html>
53. Sinclair, A., Xie, X., Teltscher, M. & Dendukuri, N. Systematic Review and Meta-Analysis of a Urine-Based Pneumococcal Antigen Test for Diagnosis of Community-Acquired Pneumonia Caused by Streptococcus pneumoniae. *J. Clin. Microbiol.* **51,** 2303–2310 (2013).
54. Dowell, S. F., Garman, R. L., Liu, G., Levine, O. S. & Yang, Y.-H. Evaluation of Binax NOW, an Assay for the Detection of Pneumococcal Antigen in Urine Samples, Performed among Pediatric Patients. *Clin. Infect. Dis.* **32,** 824–825 (2001).
55. Domínguez, J. *et al.* Usefulness of urinary antigen detection by an immunochromatographic test for diagnosis of pneumococcal pneumonia in children. *J.*

Clin. Microbiol. **41,** 2161–3 (2003).
56. Korppi, M., Leinonen, M. & Ruuskanen, O. Pneumococcal serology in children's respiratory infections. *Eur. J. Clin. Microbiol. Infect. Dis.* **27,** 167–175 (2008).
57. Watt, J. P. *et al.* Use of serology and urine antigen detection to estimate the proportion of adult community-acquired pneumonia attributable to Streptococcus pneumoniae. *Epidemiol. Infect.* **138,** 1796–803 (2010).
58. Scott, J. A. G. *et al.* Diagnosis of invasive pneumococcal disease among children in Kenya with enzyme-linked immunosorbent assay for immunoglobulin G antibodies to pneumococcal surface adhesin A. *Clin. Diagn. Lab. Immunol.* **12,** 1195–201 (2005).
59. Albrich, W. C. *et al.* Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. *Clin Infect Dis* **54,** 601–609 (2012).
60. Chochua, S. *et al.* Increased Nasopharyngeal Density and Concurrent Carriage of Streptococcus pneumoniae, Haemophilus influenzae, and Moraxella catarrhalis Are Associated with Pneumonia in Febrile Children. doi:10.1371/journal.pone.0167725
61. Wright, A. E., Parry Morgan, W., Colebrook, L. & Dodgson, R. W. Observations on prophylactic inoculation against pneumococcus infections, and on the results which have been achieved by it. *Lancet* **183,** 87–95 (1914).
62. Heffron, R. Pneumonia with special reference to pneumococcal lobar pneumonia. *New York, NY Commonw. Fund* 446–483, 805–936, 946 (1939).
63. Tillett, W. S. & Francis, T. Serological reactions in pneumonia with a non-protein somatic fraction of pneumococcus. *J. Exp. Med.* **52,** 561–71 (1930).
64. MacLeod, C. M. & Hodges, R. G. Prevention of pneumococcal pneumonia by immunization with specific capsular polysaccharides. *J. Exp. Med.* **82,** 445–65 (1945).
65. Black, S. *et al.* Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr. Infect. Dis. J.* **19,** 187–95 (2000).
66. Safety and Immunogenicity of a 10 Valent Pneumococcal Conjugate Vaccine (SIILPCV10) in Healthy Adults, Toddlers, Infants. at <https://clinicaltrials.gov/ct2/show/NCT02308540>
67. McFetridge, R. *et al.* Safety, tolerability, and immunogenicity of 15-valent pneumococcal conjugate vaccine in healthy adults. *Vaccine* **33,** 2793–2799 (2015).
68. Morbidity and Mortality Weekly Report. at <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5909a2.htm>
69. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)* **385,** 117–71 (2015).
70. O'Brien, K. L. *et al.* Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet* **374,** 893–902 (2009).
71. Lucero, M. G. *et al.* Efficacy of an 11-valent pneumococcal conjugate vaccine against radiologically confirmed pneumonia among children less than 2 years of age in the Philippines: a randomized, double-blind, placebo-controlled trial. *Pediatr Infect Dis J* **28,** 455–462 (2009).
72. Cutts, F. T. *et al.* Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *Lancet* **365,** 1139–1146 (2005).
73. Klugman, K. P. *et al.* A trial of a 9-valent pneumococcal conjugate vaccine in children with and those without HIV infection. *N Engl J Med* **349,** 1341–1348 (2003).
74. Hansen, J. *et al.* Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than 5 years of age for prevention of pneumonia: updated analysis

using World Health Organization standardized interpretation of chest radiographs. *Pediatr Infect Dis J* **25,** 779–781 (2006).
75. World Health Organization. *Global Invasive Bacterial and Rotavirus Surveillance Bulletin*. (2013). at <http://www.who.int/immunization/monitoring_surveillance/resources/Global_IBDRV_Surv_Bulletin_July_2015.pdf>
76. Satzke, C. *et al.* Standard method for detecting upper respiratory carriage of Streptococcus pneumoniae: Updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. *Vaccine* **32,** 165–179 (2013).
77. Miller, E., Andrews, N. J., Waight, P. A., Slack, M. P. & George, R. C. 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 (2011).
78. Pebody, R. G. *et al.* Pneumococcal vaccination policy in Europe. *Euro Surveill.* **10,** 174–8 (2005).
79. Isaacman, D. J., McIntosh, E. D. & Reinert, R. R. Burden of invasive pneumococcal disease and serotype distribution among Streptococcus pneumoniae isolates in young children in Europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines. *Int. J. Infect. Dis.* **14,** e197-209 (2010).
80. Office for National Statistics - Population Estimates for UK, England and Wales, Scotland and Northern Ireland. (2016). at <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>
81. Office for National Statistics. Population estimates. at <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates>
82. The Kings Fund. The NHS budget and how it has changed | The King's Fund. at <https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/nhs-budget>
83. Government of Nepal Central Bureau of Statistics. *National Population and Housing Census 2011*. (2011). at <http://cbs.gov.np/image/data/Population/Major Highlights/Major-Finding.pdf>
84. Central Intelligence Agency. The World Factbook — Central Intelligence Agency. at <https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2004rank.html>
85. Government of Nepal Office of the Investment Board. Health. at <http://ibn.gov.np/health>
86. Nair, H. *et al.* Pneumonia hospitalisations in Scotland following the introduction of pneumococcal conjugate vaccination in young children. *BMC Infect. Dis.* **16,** 390 (2016).
87. Trotter, C. L., Stuart, J. M., George, R. & Miller, E. Increasing hospital admissions for pneumonia, England. *Emerg. Infect. Dis.* **14,** 727–33 (2008).
88. Gessner, B. D., Mueller, J. E. & Yaro, S. African meningitis belt pneumococcal disease epidemiology indicates a need for an effective serotype 1 containing vaccine, including for older children and adults. *BMC Infect. Dis.* **10,** 22 (2010).
89. Saxena, S. *et al.* Additive impact of pneumococcal conjugate vaccines on pneumonia and empyema hospital admissions in England. *J. Infect.* **71,** 428–436 (2015).
90. Waight, P. A. *et al.* 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 (2015).

91. Hamaluba, M. *et al.* A cross-sectional observational study of pneumococcal carriage in children, their parents, and older adults following the introduction of the 7-valent pneumococcal conjugate vaccine. *Med.* **94,** e335 (2015).
92. Countries with PCV impact evaluation. at <http://view-hub.org/viz/>
93. Gray, B. M., Converse, G. M. & Dillon, H. C. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J. Infect. Dis.* **142,** 923–33 (1980).
94. Cardozo, D. M. *et al.* Prevalence and risk factors for nasopharyngeal carriage of Streptococcus pneumoniae among adolescents. *J. Med. Microbiol.* **57,** 185–189 (2008).
95. Mackenzie, G. A., Leach, A. J., Carapetis, J. R., Fisher, J. & Morris, P. S. Epidemiology of nasopharyngeal carriage of respiratory bacterial pathogens in children and adults: cross-sectional surveys in a population with high rates of pneumococcal disease. *BMC Infect. Dis.* **10,** 304 (2010).
96. Principi, N., Marchisio, P., Schito, G. C. & Mannelli, S. Risk factors for carriage of respiratory pathogens in the nasopharynx of healthy children. Ascanius Project Collaborative Group. *Pediatr. Infect. Dis. J.* **18,** 517–23 (1999).
97. Sleeman, K. L. *et al.* Capsular serotype-specific attack rates and duration of carriage of Streptococcus pneumoniae in a population of children. *J. Infect. Dis.* **194,** 682–8 (2006).
98. Principi, N., Marchisio, P., Schito, G. C. & Mannelli, S. Risk factors for carriage of respiratory pathogens in the nasopharynx of healthy children. Ascanius Project Collaborative Group. *Pediatr. Infect. Dis. J.* **18,** 517–23 (1999).
99. Turner, P. *et al.* A longitudinal study of Streptococcus pneumoniae carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. *PLoS One* **7,** e38271 (2012).
100. Pebody, R. G. *et al.* Use of antibiotics and risk factors for carriage of Streptococcus pneumoniae: a longitudinal household study in the United Kingdom. *Epidemiol. Infect.* **137,** 555 (2009).
101. Varon, E. *et al.* Impact of antimicrobial therapy on nasopharyngeal carriage of Streptococcus pneumoniae, Haemophilus influenzae, and Branhamella catarrhalis in children with respiratory tract infections. *Clin. Infect. Dis.* **31,** 477–81 (2000).
102. Gray, B. M., Converse, G. M. & Dillon, H. C. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J. Infect. Dis.* **142,** 923–33 (1980).
103. Flasche, S. *et al.* Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in England: a cross-sectional study. *PLoS Med* **8,** e1001017 (2011).
104. Tocheva, A. S., Jefferies, J. M. C., Christodoulides, M., Faust, S. N. & Clarke, S. C. 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* **31,** 3187–3190 (2013).
105. Flasche, S. *et al.* The impact of specific and non-specific immunity on the ecology of Streptococcus pneumoniae and the implications for vaccination. *Proc Biol Sci* **280,** 20131939 (2013).
106. *Management of community acquired pneumonia in childhood 2002*. (2002). at <https://www.brit-thoracic.org.uk/document-library/clinical-information/pneumonia/paediatric-pneumonia/management-of-community-acquired-pneumonia-in-childhood-2002/management-of-community-acquired-pneumonia-in-childhood-2002/>

107. Harris, M. *et al.* Guidelines for the management of community acquired pneumonia in children: update 2011 British Thoracic Society Community Acquired Pneumonia in Children Guideline Group. *Thorax* **66,** (2011).
108. *Prescriptions Dispensed in the Community*. (2015). at <http://content.digital.nhs.uk/catalogue/PUB17644/pres-disp-com-eng-2004-14-rep.pdf>
109. *National Antibiotic Treatment Guidelines*. (2014). at <http://www.mohp.gov.np/images/pdf/guideline/National_Antibiotic_Treatment_Guidelines.pdf>
110. *Revised WHO classification and treatment of childhood pneumonia at health facilities*. at <http://apps.who.int/iris/bitstream/10665/137319/1/9789241507813_eng.pdf>
111. Wachter, D. A., Joshi, M. P. & Rimal, B. Antibiotic dispensing by drug retailers in Kathmandu, Nepal. *Trop. Med. Int. Heal.* **4,** 782–788 (1999).
112. Van Boeckel, T. P. *et al.* Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *Lancet. Infect. Dis.* **14,** 742–50 (2014).
113. Hansman, D. & Bullen, M. M. A Resistant Pneumococcus. *Lancet* **290,** 264–265 (1967).
114. Kislak, J. W., Razavi, L. M., Daly, A. K. & Finland, M. Susceptibility of pneumococci to nine antibiotics. *Am. J. Med. Sci.* **250,** 261–8 (1965).
115. *DH UK 5 Year Antimicrobial Resistance (AMR) Strategy 2013-2018*. at <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/553496/2nd_UK_AMR_annual_report.pdf>
116. Khanal, B. *et al.* Antimicrobial Resistant Streptococcus pneumoniae. *J Nepal Med Assoc* **49,** 220–4 (2010).
117. Joshi, H. H., Gertz, R. E., da Gloria Carvalho, M. & Beall, B. W. Use of Silica Desiccant Packets for Specimen Storage and Transport To Evaluate Pneumococcal Nasopharyngeal Carriage among Nepalese Children. *J. Clin. Microbiol.* **46,** 3175–3176 (2008).
118. Easow, J. M. *et al.* Streptococcus pneumoniae infections in western Nepal. *Southeast Asian J. Trop. Med. Public Health* **42,** 912–9 (2011).
119. Shah, A. S. *et al.* Invasive Pneumococcal Disease in Kanti Children's Hospital, Nepal, as Observed by the South Asian Pneumococcal Alliance Network. *Clin. Infect. Dis.* **48,** S123–S128 (2009).
120. Hall, L. M. Application of molecular typing to the epidemiology of Streptococcus pneumoniae. *J. Clin. Pathol.* **51,** 270–4 (1998).
121. Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science* **293,** 498–506 (2001).
122. Devine, V. T. *et al.* The rise and fall of pneumococcal serotypes carried in the PCV era. *Vaccine* **35,** 1293–1298 (2017).
123. Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat. Genet.* **45,** 656–63 (2013).
124. Chewapreecha, C. *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat. Genet.* **46,** 305–9 (2014).
125. Everett, D. B. *et al.* Genetic characterisation of Malawian pneumococci prior to the roll-out of the PCV13 vaccine using a high-throughput whole genome sequencing approach. *PLoS One* **7,** e44250 (2012).
126. Planned changes to the routine Childhood Immunisation Programme. (2006). at <http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Lettersandcirculars/Dearcolleagueletters/DH_4128120>
127. Andrews, N. *et al.* Using the indirect cohort design to estimate the effectiveness of the

seven valent pneumococcal conjugate vaccine in England and Wales. *PLoS One* **6,** e28435 (2011).
128. van Hoek, A. J. *et al.* Pneumococcal carriage in children and adults two years after introduction of the thirteen valent pneumococcal conjugate vaccine in England. *Vaccine* **32,** 4349–4355 (2014).
129. Dagan, R. *et al.* Comparative immunogenicity and efficacy of 13-valent and 7-valent pneumococcal conjugate vaccines in reducing nasopharyngeal colonization: a randomized double-blind trial. *Clin. Infect. Dis.* **57,** 952–62 (2013).
130. Demczuk, W. H. B. *et al.* Serotype distribution of invasive Streptococcus pneumoniae in Canada after the introduction of the 13-valent pneumococcal conjugate vaccine, 2010-2012. *Can. J. Microbiol.* **59,** 778–88 (2013).
131. Fry, N. *et al.* The fall and rise of serotype 19A in invasive pneumococcal disease: application of whole genome sequencing to investigate the recent rise in England and Wales. in *The 10th The International Symposium on Pneumococci and Pneumococcal Diseases* (2016). at <https://ep70.eventpilot.us/web/page.php?page=Inthtml&project=ISPPD16&id=abstract_199615>
132. Satzke, C. *et al.* Standard method for detecting upper respiratory carriage of Streptococcus pneumoniae: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. *Vaccine* **32,** 165–179 (2013).
133. Elberse, K. E. M. *et al.* Seroprevalence of IgG antibodies against 13 vaccine Streptococcus pneumoniae serotypes in the Netherlands. *Vaccine* **29,** 1029–35 (2011).
134. R Core Team (2016). R: A language and environment for statistical computing. at <https://www.r-project.org/>
135. Lee, G. M. *et al.* Impact of 13-Valent Pneumococcal Conjugate Vaccination on Streptococcus pneumoniae Carriage in Young Children in Massachusetts. *J. Pediatric Infect. Dis. Soc.* **3,** 23–32 (2014).
136. Grant, L. R. *et al.* Comparative immunogenicity of 7 and 13-valent pneumococcal conjugate vaccines and the development of functional antibodies to cross-reactive serotypes. *PLoS One* **8,** e74906 (2013).
137. Baby Vaccine Study (Sched3). at <https://clinicaltrials.gov/ct2/show/NCT02482636>
138. Pace, D. *et al.* Immunogenicity of reduced dose priming schedules of serogroup C meningococcal conjugate vaccine followed by booster at 12 months in infants: open label randomised controlled trial. *BMJ* **350,** h1554 (2015).
139. Cutts, F. T. *et al.* Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *Lancet (London, England)* **365,** 1139–46 (2005).
140. Lucero, M. G. *et al.* Efficacy of an 11-valent pneumococcal conjugate vaccine against radiologically confirmed pneumonia among children less than 2 years of age in the Philippines: a randomized, double-blind, placebo-controlled trial. *Pediatr. Infect. Dis. J.* **28,** 455–62 (2009).
141. O'Brien, K. L., Nohynek, H. & World Health Organization Pneumococcal Vaccine Trials Carraige Working, G. Report from a WHO working group: standard method for detecting upper respiratory carriage of Streptococcus pneumoniae. *Pediatr Infect Dis J* **22,** 133–140 (2003).
142. Satzke, C. *et al.* The PneuCarriage Project: A Multi-Centre Comparative Study to Identify the Best Serotyping Methods for Examining Pneumococcal Carriage in Vaccine Evaluation Studies. *PLOS Med.* **12,** e1001903 (2015).
143. Ussery, X. T. *et al.* Risk factors for nasopharyngeal carriage of resistant Streptococcus pneumoniae and detection of a multiply resistant clone among children living in the

Yukon-Kuskokwim Delta region of Alaska. *Pediatr. Infect. Dis. J.* **15,** 986–92 (1996).
144. López, B. *et al.* Epidemiological study of Streptococcus pneumoniae carriers in healthy primary-school children. *Eur. J. Clin. Microbiol. Infect. Dis.* **18,** 771–6 (1999).
145. Huebner, R. E., Dagan, R., Porath, N., Wasas, A. D. & Klugman, K. P. Lack of utility of serotyping multiple colonies for detection of simultaneous nasopharyngeal carriage of different pneumococcal serotypes. *Pediatr. Infect. Dis. J.* **19,** 1017–20 (2000).
146. Turner, P. *et al.* Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. *J. Clin. Microbiol.* **49,** 1784–9 (2011).
147. Kandasamy, R. *et al.* Multi-serotype pneumococcal nasopharyngeal carriage prevalence in vaccine naïve Nepalese children, assessed using molecular serotyping. *PLoS One* **10,** e0114286 (2015).
148. Kamng'ona, A. W. *et al.* High multiple carriage and emergence of Streptococcus pneumoniae vaccine serotype variants in Malawian children. *BMC Infect. Dis.* **15,** 234 (2015).
149. Rodrigues, F. *et al.* Pneumococcal Serotypes Colonise the Nasopharynx in Children at Different Densities. *PLoS One* **11,** e0163435 (2016).
150. Kandasamy, R. *et al.* Multi-serotype pneumococcal nasopharyngeal carriage prevalence in vaccine naive nepalese children, assessed using molecular serotyping. *PLoS One* **10,** e0114286 (2015).
151. Turner, P. *et al.* Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. *J. Clin. Microbiol.* **49,** 1784–9 (2011).
152. Buckingham, S. C., King, M. D. & Miller, M. L. Incidence and etiologies of complicated parapneumonic effusions in children, 1996 to 2001. *Pediatr. Infect. Dis. J.* **22,** 499–503 (2003).
153. Schaffner, A., Michel-Harder, C. & Yeginsoy, S. Detection of capsular polysaccharide in serum for the diagnosis of pneumococcal pneumonia: clinical and experimental evaluation. *J. Infect. Dis.* **163,** 1094–102 (1991).
154. Tharpe, J. A. *et al.* Comparison of a pneumococcal common protein (PsaA) antibody ELISA and a PsaA immune complex ELISA for detection of pneumococcal serum antibody. *Pathobiology* **66,** 77–83 (1998).
155. Scott, J. A. G., Obiero, J., Hall, A. J. & Marsh, K. Validation of Immunoglobulin G Enzyme-Linked Immunosorbent Assay for Antibodies to Pneumococcal Surface Adhesin A in the Diagnosis of Pneumococcal Pneumonia among Adults in Kenya. *J. Infect. Dis.* **186,** 220–226 (2002).
156. Tuerlinckx, D. *et al.* Evaluation of a WHO-validated Serotype-specific Serological Assay for the Diagnosis of Pneumococcal Etiology in Children With Community-acquired Pneumonia. *Pediatr. Infect. Dis. J.* **32,** e277–e284 (2013).
157. De Schutter, I. *et al.* Pneumococcal Aetiology and Serotype Distribution in Paediatric Community-Acquired Pneumonia. *PLoS One* **9,** e89013 (2014).
158. Baggett, H. C. *et al.* Density of Upper Respiratory Colonization With Streptococcus pneumoniae and Its Role in the Diagnosis of Pneumococcal Pneumonia Among Children Aged <5 years in the PERCH Study. *Clin. Infect. Dis.* **64,** S317–S327 (2017).
159. Albrich, W. C. *et al.* Genomic Load from Sputum Samples and Nasopharyngeal Swabs for Diagnosis of Pneumococcal Pneumonia in HIV-Infected Adults. *J. Clin. Microbiol.* **52,** 4224–4229 (2014).
160. Stralin, K. *et al.* Comparison of Sputum and Nasopharyngeal Aspirate Samples and of the PCR Gene Targets lytA and Spn9802 for Quantitative PCR for Rapid Detection of Pneumococcal Pneumonia. *J. Clin. Microbiol.* **52,** 83–89 (2014).

161. Carvalho, M. da G. S. *et al.* Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. *J. Clin. Microbiol.* **45,** 2460–6 (2007).
162. StepOne™ and StepOnePlus™ Real-Time PCR System. at <https://www.thermofisher.com/uk/en/home/technical-resources/software-downloads/StepOne-and-StepOnePlus-Real-Time-PCR-System.html>
163. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12,** 77 (2011).
164. Chewapreecha, C. *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* **46,** 305–309 (2014).
165. Cleary, D. W. *et al.* Comparative Genomics of Carriage and Disease Isolates of Streptococcus pneumoniae Serotype 22F Reveals Lineage-Specific Divergence and Niche Adaptation. *Genome Biol. Evol.* **8,** 1243–51 (2016).
166. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15,** R46 (2014).
167. SMALT. at <http://www.sanger.ac.uk/science/tools/smalt-0>
168. Keane, J. A. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genomics* **2,** (2016).
169. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–3 (2014).
170. Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30,** 1224–8 (2013).
171. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–4 (2010).
172. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* **14,** 1394–1403 (2004).
173. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16,** 944–5 (2000).
174. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43,** e15–e15 (2014).
175. Salter, S. J. *et al.* Variation at the capsule locus, cps, of mistyped and non-typable Streptococcus pneumoniae isolates. *Microbiology* **158,** 1560–9 (2012).
176. Iannelli, F., Oggioni, M. R. & Pozzi, G. Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae. *Gene* **284,** 63–71 (2002).
177. Hollingshead, S. K., Becker, R. & Briles, D. E. Diversity of PspA: mosaic genes and evidence for past recombination in Streptococcus pneumoniae. *Infect. Immun.* **68,** 5889–900 (2000).
178. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–7 (2004).
179. Iqbal, Z., Caccamo, M., Turner, I., Flicek, P. & McVean, G. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.* **44,** 226–32 (2012).
180. Croucher, N. J. *et al.* Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat. Commun.* **5,** 5471 (2014).
181. Hollingshead, S. K., Becker, R. & Briles, D. E. Diversity of PspA: mosaic genes and evidence for past recombination in Streptococcus pneumoniae. *Infect. Immun.* **68,** 5889–900 (2000).

182. Hilty, M. *et al.* Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. *Genome Biol Evol* **6,** 3281–3294 (2014).
183. Kallio, A. *et al.* Role of Pht proteins in attachment of Streptococcus pneumoniae to respiratory epithelial cells. *Infect. Immun.* **82,** 1683–91 (2014).
184. Lees, J. A. *et al.* Large scale genomic analysis shows no evidence for pathogen adaptation between the blood and cerebrospinal fluid niches during bacterial meningitis. *Microb. genomics* **3,** e000103 (2017).
185. Malla, S., Kansakar, P., Serichantalergs, O., Rahman, M. & Basnet, S. Epidemiology of typhoid and paratyphoid fever in Kathmandu: two years study and trends of antimicrobial resistance. *JNMA. J. Nepal Med. Assoc.* **44,** 18–22
186. Fernández, L. & Hancock, R. E. W. Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* **25,** 661–81 (2012).
187. Daly, M. M., Doktor, S., Flamm, R. & Shortridge, D. Characterization and prevalence of MefA, MefE, and the associated msr(D) gene in Streptococcus pneumoniae clinical isolates. *J. Clin. Microbiol.* **42,** 3570–4 (2004).
188. Roberts, M. C. *et al.* Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. *Antimicrob. Agents Chemother.* **43,** 2823–30 (1999).
189. Abraham, E. P. & Chain, E. An enzyme from bacteria able to destroy penicillin. 1940. *Rev. Infect. Dis.* **10,** 677–8
190. Coffey, T. J., Dowson, C. G., Daniels, M. & Spratt, B. G. Genetics and Molecular Biology of β-Lactam-Resistant Pneumococci. *Microb. Drug Resist.* **1,** 29–34 (1995).
191. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6,** 90 (2014).
192. Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58,** 212–20 (2014).
193. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44,** W242–W245 (2016).
194. Schwarz, S., Kehrenberg, C., Doublet, B. & Cloeckaert, A. Molecular basis of bacterial resistance to chloramphenicol and florfenicol. *FEMS Microbiol. Rev.* **28,** 519–542 (2004).
195. Connell, S. R., Tracz, D. M., Nierhaus, K. H. & Taylor, D. E. Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrob. Agents Chemother.* **47,** 3675–81 (2003).
196. Leclercq, R. Mechanisms of Resistance to Macrolides and Lincosamides: Nature of the Resistance Elements and Their Clinical Implications. *Clin. Infect. Dis.* **34,** 482–492 (2002).
197. Moore, C. E. *et al.* Characterisation of Invasive Streptococcus pneumoniae Isolated from Cambodian Children between 2007 – 2012. *PLoS One* **11,** e0159358 (2016).
198. Weinberger, D. M., Malley, R. & Lipsitch, M. Serotype replacement in disease after pneumococcal vaccination. *Lancet (London, England)* **378,** 1962–73 (2011).
199. Johnson, H. L. *et al.* Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Med.* **7,** (2010).
200. Enright, M. C. & Spratt, B. G. Multilocus sequence typing. *Trends Microbiol.* **7,** 482–7 (1999).
201. Clarke, S. C., Scott, K. J. & McChlery, S. M. Serotypes and sequence types of

pneumococci causing invasive disease in Scotland prior to the introduction of pneumococcal conjugate polysaccharide vaccines. *J. Clin. Microbiol.* **42,** 4449–52 (2004).
202. Wyres, K. L. *et al.* Pneumococcal capsular switching: a historical perspective. *J. Infect. Dis.* **207,** 439–49 (2013).
203. The Global Pneumococcal Sequencing Project. at <http://www.pneumogen.net/gps/index.html>
204. Realtime PCR oligonucleotides. at <https://www.cdc.gov/streplab/downloads/realtime-pcr-oligonucleotides.pdf>
205. Ko, K. S., Baek, J. Y. & Song, J.-H. Capsular gene sequences and genotypes of "serotype 6E" Streptococcus pneumoniae isolates. *J. Clin. Microbiol.* **51,** 3395–9 (2013).
206. Multilocus sequence typing by blast using the schemes from PubMLST.
207. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11,** 595 (2010).
208. Inouye, M., Conway, T. C., Zobel, J. & Holt, K. E. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC Genomics* **13,** 338 (2012).
209. Nascimento, M. *et al.* PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* **33,** 128–129 (2017).
210. Francisco, A. P., Bugalho, M., Ramirez, M. & Carriço, J. A. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10,** 152 (2009).
211. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).
212. CARTO - Location Intelligence Software. at <https://carto.com/>
213. Marimón, J. M., Ercibengoa, M., Tamayo, E., Alonso, M. & Pérez-Trallero, E. Long-Term Epidemiology of Streptococcus pneumoniae Serogroup 6 in a Region of Southern Europe with Special Reference to Serotype 6E. *PLoS One* **11,** e0149047 (2016).
214. Baby Vaccine Study (Sched3). at <https://clinicaltrials.gov/ct2/show/NCT02482636?term=1%2B1&rank=7>
215. Deloria Knoll, M. *et al.* Systematic review of the effect of pneumococcal conjugate vaccine dosing schedules on immunogenicity. *Pediatr. Infect. Dis. J.* **33 Suppl 2,** S119-29 (2014).

Appendix A

Table 38. ***Serotype-specific NVT pneumococcal carriage prevalence among UK children.***

Serotype	Sleeman 2006, n (%)	Meats 2003, n (%)	Tocheva 2011 (2006-2007), n (%)	Tocheva 2011 (2007-2008), n (%)	Tocheva 2011 (200-2009), n (%)	Devine 2017 (2009-2010), n (%)	Devine 2017 (2010-2011), n (%)	Hamalub a 2015, n (%)	Flasch e 2011, n (%)	Devine 2017 (2011-2012), n (%)	van Hoek 2014, n (%)	Devine 2017 (2012-2013), n (%)
11A	22 (1.2)	5 (0.4)	3 (1)	9 (2.4)	7 (2.13)	13 (3.3)	7 (2.4)	24 (4.2)	4 (2.1)	12 (3.6)	18 (6.5)	5 (2.2)
23B	-	1 (0.1)	1 (0.3)	1 (0.3)	5 (1.52)	6 (1.5)	8 (2.8)	24 (4.2)	6 (3.1)	11 (3.3)	13 (4.7)	7 (3.1)
24F	-	2 (0.2)	-	-	-	1 (0.3)	-	1 (0.2)	2 (1)	1 (0.3)	9 (3.3)	4 (1.8)
35F	11 (0.6)	1 (0.1)	1 (0.3)	3 (0.8)	2 (0.61)	7 (1.8)	9 (3.1)	14 (2.4)	1 (0.5)	4 (1.2)	10 (3.6)	2 (0.9)
6C	-	-	4 (1.2)	14 (3.8)	14 (4.27)	4 (1)	9 (3.1)	25 (4.4)	8 (4.2)	6 (1.8)	7 (2.5)	4 (1.8)
16F	7 (0.4)	3 (0.2)	2 (0.6)	2 (0.5)	-	1 (0.3)	4 (1.4)	8 (1.4)	-	1 (0.3)	9 (3.3)	3 (1.3)
21	9 (0.5)	5 (0.4)	1 (0.3)	2 (0.5)	2 (0.61)	10 (2.5)	13 (4.5)	12 (2.1)	5 (2.6)	4 (1.2)	9 (3.3)	3 (1.3)
23A	17 (0.9)	4 (0.3)	1 (0.3)	3 (0.8)	5 (1.52)	2 (0.5)	4 (1.4)	12 (2.1)	4 (2.1)	6 (1.8)	6 (2.2)	2 (0.9)
15B	13 (0.7)	-	5 (1.5)	-	2 (0.61)	4 (1)	7 (2.4)	2 (0.4)	6 (3.1)	5 (1.5)	7 (2.5)	10 (4.5)
22F	18 (1)	4 (0.3)	2 (0.6)	5 (1.3)	9 (2.74)	9 (2.3)	3 (1)	13 (2.3)	3 (1.6)	2 (0.6)	8 (2.9)	2 (0.9)
10A	28 (1.5)	2 (0.2)	-	-	3 (0.91)	-	1 (0.3)	10 (1.7)	2 (1)	6 (1.8)	5 (1.8)	4 (1.8)
17F	-	4 (0.3)	-	-	-	2 (0.5)	-	1 (0.2)	-	1 (0.3)	5 (1.8)	-
15C	9 (0.5)	-	2 (0.6)	1 (0.3)	-	5 (1.3)	2 (0.7)	8 (1.4)	4 (2.1)	5 (1.5)	3 (1.1)	2 (0.9)
35A	-	-	-	-	-	-	-	-	-	-	4 (1.4)	-
35B	-	-	-	-	2 (0.61)	-	3 (1)	12 (2.1)	-	5 (1.5)	3 (1.1)	7 (3.1)
15A	-	2 (0.15)	1 (0.3)	-	2 (0.61)	4 (1)	6 (2.1)	4 (0.7)	-	10 (3)	2 (0.7)	10 (4.5)
31	-	-	2 (0.6)	-	1 (0.3)	-	2 (0.7)	-	2 (1)	3 (0.9)	2 (0.7)	1 (0.4)

38	-	3 (0.22)	2 (0.6)	-	2 (0.61)	2 (0.5)	1 (0.3)	4 (0.7)	-	1 (0.3)	2 (0.7)	1 (0.4)
11D	-	-	-	-	-	-	-	1 (0.2)	-	-	1 (0.4)	-
15	-	-	-	-	-	-	-	-	-	-	1 (0.4)	-
23C	-	-	-	-	-	-	-	-	-	-	1 (0.4)	-
33F	8 (0.4)	4 (0.3)	-	1 (0.3)	3 (0.91)	4 (1)	8 (2.8)	7 (1.2)	4 (2.1)	1 (0.3)	1 (0.4)	3 (1.3)
35C	-	-	-	-	-	-	-	-	-	-	1 (0.4)	-
7C	-	-	-	1 (0.3)	1 (0.3)	-	-	-	-	-	1 (0.4)	-
15B/C#	-	12 (0.9)	-	-	-	-	-	-	-	-	-	-
9N	16 (0.8)	5 (0.4)	1 (0.3)	-	1 (0.3)	2 (0.5)	-	-	1 (0.5)	-	-	-
8	7 (0.4)	4 (0.3)	-	-	-	1 (0.3)	-	-	-		-	-
20	2 (0.1)	3 (0.2)	-	-	-	1 (0.3)	1 (0.3)	1 (0.2)	-	-	-	-
12F	-	2 (0.2)	-	-	-	-	1 (0.3)	2 (0.4)	-	-	-	1 (0.4)
13	-	1 (0.1)	-	-	-	-	-	1 (0.2)	-	-	-	-
27	-	1 (0.1)	-	-	-	-	-	-	-	-	-	-
31	1 (0.1)	1 (0.1)	-	-	-	-	-	7 (1.2)	-	-	-	-
37	-	1 (0.1)	-	-	-	-	-	-	1 (0.5)	-	-	-
NT	-	2 (0.2)	1 (0.3)	-	4 (1.2)	2 (0.5)	-	9 (1.6)	-	-	-	1 (0.4)
11C	-	-	-	-	-	-	-	-	6 (3.1)	-	-	-
11B	-	-	-	-	-	-	-	-	3 (1.6)	-	-	-
33A	-	-	-	-	-	-	-	-	2 (1)	-	-	-
17A	-	-	-	-	-	-	-	-	1 (0.5)	-	-	-
22A	-	-	-	-	-	-	-	1 (0.2)	1 (0.5)	-	-	-
28F	-	-	-	-	-	-	-	-	1 (0.5)	-	-	-
29	-	-	-	-	-	-	-	6 (1)	1 (0.5)	-	-	-
22C	-	-	-	-	-	-	-	1 (0.2)	-	-	-	-
24A	-	-	-	-	-	-	-	1 (0.2)	-	-	-	-
34	-	-	1 (0.3)	1 (0.3)	-	2 (0.5)	-	-	-	2 (0.6)	-	1 (0.4)
33D	-	-	-	-	1 (0.3)	-	-	-	-	-	-	-

6D	-	-	-	1 (0.3)	-	-	-	-	-	-	-	-
6F	-	-	-	-	-	1 (0.3)	-	-	-	-	-	-
Unassigned	-	-	-	-	-	-	-	-	-	5 (1.5)	-	2 (0.9)

Table 39. ***Serotype-specific NVT pneumococcal carriage prevalence among Nepalese children.***

Serotype	Hanieh (2014) rural SDP, n (%) N=526	Hanieh (2014) urban SDP, n (%) N=574	Hanieh (2014) urban STGG, n (%) N=574	Kandasamy (2015) urban STGG, n (%) N=600
NT	17 (3.2)	27 (4.7)	28 (4.9)	23 (3.8)
6^	19 (3.6)	19 (3.3)	48 (8.4)	-
15B/C	12 (2.3)	11 (1.9)	24 (4.2)	-
34	6 (1.1)	12 (2.1)	16 (2.8)	8 (1.3)
11A	11 (2.1)	6 (1)	9 (1.6)	1 (0.2)
23A	3 (0.6)	6 (1)	12 (2.1)	4 (0.7)
15B	1 (0.2)	3 (0.5)	3 (0.5)	17 (2.8)
17F	1 (0.2)	4 (0.7)	12 (2.1)	7 (1.2)
21	-	5 (0.9)	13 (2.3)	4 (0.7)
35B	-	6 (1)	9 (1.6)	7 (1.2)
16F	3 (0.6)	4 (0.7)	6 (1)	6 (1)
10A	1 (0.2)	3 (0.5)	4 (0.7)	10 (1.7)
35A	2 (0.4)	1 (0.2)	4 (0.7)	10 (1.7)
23B	2 (0.4)	4 (0.7)	9 (1.6)	2 (0.3)
13	5 (1)	2 (0.3)	6 (1)	3 (0.5)
7C	1 (0.2)	2 (0.3)	7 (1.2)	4 (0.7)
20	5 (1)	1 (0.2)	3 (0.5)	5 (0.8)
31	1 (0.2)	4 (0.7)	7 (1.2)	2 (0.3)
39	4 (0.8)	2 (0.3)	2 (0.3)	5 (0.8)
15A	2 (0.4)	5 (0.9)	4 (0.7)	2 (0.3)
33B	1 (0.2)	3 (0.5)	5 (0.9)	4 (0.7)
33F	6 (1.1)	2 (0.3)	4 (0.7)	1 (0.2)
9N/L	4 (0.8)	2 (0.3)	6 (1)	-
18	7 (1.3)	-	4 (0.7)	-
35F	-	3 (0.5)	-	8 (1.3)
35F/47F	3 (0.6)	2 (0.3)	5 (0.9)	-
15C	1 (0.2)	5 (0.9)	1 (0.2)	3 (0.5)
11D	-	-	-	8 (1.3)
24A/B/F	3 (0.6)	-	5 (0.9)	-
19B	2 (0.4)	3 (0.5)	2 (0.3)	1 (0.2)
6C	-	1 (0.2)	-	6 (1)
8	2 (0.4)	-	1 (0.2)	3 (0.5)
22F	-	2 (0.3)	-	4 (0.7)
22F/A	3 (0.6)	-	3 (0.5)	-
35C	-	1 (0.2)	-	5 (0.8)
45	1 (0.2)	-	-	4 (0.7)
10F	1 (0.2)	2 (0.3)	1 (0.2)	1 (0.2)
11B	-	2 (0.3)	2 (0.3)	1 (0.2)
28F	-	3 (0.5)	-	2 (0.3)
7B	-	2 (0.3)	-	3 (0.5)

9N	-	2 (0.3)	-	3 (0.5)
9L	-	3 (0.5)	-	1 (0.2)
9V/9A[#]	8 (1.5)	4 (0.7)	5 (0.9)	-
12F	1 (0.2)	1 (0.2)	1 (0.2)	1 (0.2)
2	-	-	3 (0.5)	-
29	-	1 (0.2)	-	2 (0.3)
36	-	1 (0.2)	1 (0.2)	1 (0.2)
17A	-	1 (0.2)	1 (0.2)	1 (0.2)
22A	-	-	1 (0.2)	2 (0.3)
33C	-	1 (0.2)	-	2 (0.3)
48	-	-	1 (0.2)	1 (0.2)
18A	-	1 (0.2)	-	1 (0.2)
19C	2 (0.4)	-	-	-
24A	-	1 (0.2)	-	1 (0.2)
24F	-	2 (0.3)	-	-
28A	-	-	2 (0.3)	-
33A	-	-	1 (0.2)	1 (0.2)
6D	-	-	-	2 (0.3)
11A/D	1 (0.2)	-	-	-

*Serotypes 7A, 10C, 35B, 32F, 24B, 15C, 38, 46, were only found on one occasion in one of the study cohorts.
^Not all serotypes within serogroup 6 were reported in Hanieh *et al*.

Appendix B

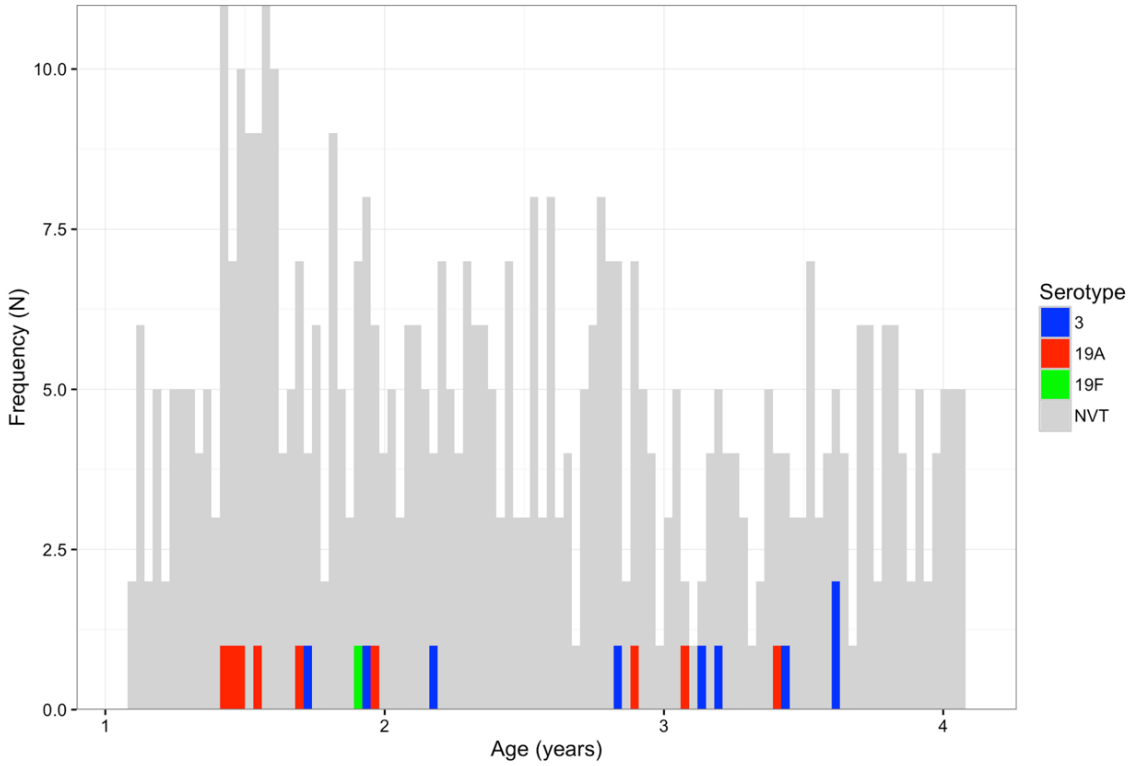


Figure 39. ***Age of UK children vaccinated with three doses of PCV13 who were found to be carrying pneumococcus.***
Enrolled children aged between 13 and 48 months were categorised according to whether they were carrying a vaccine covered serotype or a NVT. Serotype 3 carriers were significantly older (mean age of 3.1 years) compared to NVT carriers (mean age of 2.5 years) (p=0.0446, Welch Two Sample t-test).

Table 40. ***Geometric mean ratios of anti-pneumococcal polysaccharide capsule IgG from serotype 3 carriers compared with non-carriers and NVT carriers.***

Serotype	Serotype 3 carriers vs non-carriers, GMR (95% CI)	p	Serotype 3 carriers vs non vaccine-type carriers, GMR (95% CI)	p
1	0.52 (0.27-1.03)	0.0607	0.44 (0.22-0.87)	**0.0202**
3	6.17 (1.96-19.47)	**0.0022**	5.98 (1.86-19.16)	**0.003**
4	0.48 (0.22-1.01)	0.0554	0.44 (0.21-0.95)	**0.0388**
5	0.59 (0.31-1.13)	0.1154	0.53 (0.28-1.04)	0.0668
6A	0.74 (0.27-2.03)	0.558	0.66 (0.24-1.85)	0.4318
6B	0.96 (0.37-2.47)	0.9307	0.8 (0.31-2.08)	0.6469
7F	0.61 (0.33-1.11)	0.107	0.6 (0.33-1.12)	0.109
9V	1.08 (0.48-2.44)	0.8486	1.1 (0.48-2.51)	0.8233
14	0.58 (0.25-1.37)	0.2175	0.56 (0.24-1.33)	0.1901
18C	0.92 (0.42-2.01)	0.8275	0.84 (0.38-1.87)	0.6722
19A	0.56 (0.19-1.61)	0.2816	0.69 (0.24-2.02)	0.5003
19F	0.31 (0.12-0.84)	**0.0226**	0.42 (0.16-1.15)	0.0932
23F	0.67 (0.23-1.95)	0.4664	0.61 (0.21-1.8)	0.3744

Table 41. ***Geometric mean ratios of anti-pneumococcal polysaccharide capsule IgG from serotype 19A carriers compared with non-carriers and NVT carriers.***

Serotype	Serotype 19A carriers vs non-carriers, GMR (95% CI)	p	Serotype 19A carriers vs non vaccine-type carriers, GMR (95% CI)	p
1	0.67 (0.35-1.26)	0.2158	0.57 (0.3-1.08)	0.085
3	0.84 (0.29-2.43)	0.7493	0.82 (0.28-2.41)	0.7212
4	0.64 (0.32-1.3)	0.2209	0.6 (0.29-1.23)	0.1637
5	0.77 (0.42-1.43)	0.4107	0.7 (0.38-1.3)	0.2646
6A	0.96 (0.37-2.49)	0.9313	0.86 (0.33-2.26)	0.7599
6B	0.55 (0.22-1.33)	0.1851	0.45 (0.18-1.12)	0.089
7F	1 (0.57-1.75)	0.9936	0.99 (0.56-1.76)	0.9846
9V	0.94 (0.44-1.98)	0.8659	0.95 (0.45-2.03)	0.9008
14	0.93 (0.42-2.06)	0.8642	0.9 (0.4-1.99)	0.7885
18C	1.05 (0.5-2.18)	0.9049	0.96 (0.46-2.02)	0.9144
19A	5.56 (2.14-14.49)	**0.0005**	6.89 (2.62-18.13)	**0.0001**
19F	2.79 (1.13-6.84)	**0.0266**	3.76 (1.51-9.31)	**0.0048**
23F	0.41 (0.15-1.13)	0.0859	0.37 (0.13-1.04)	0.0608

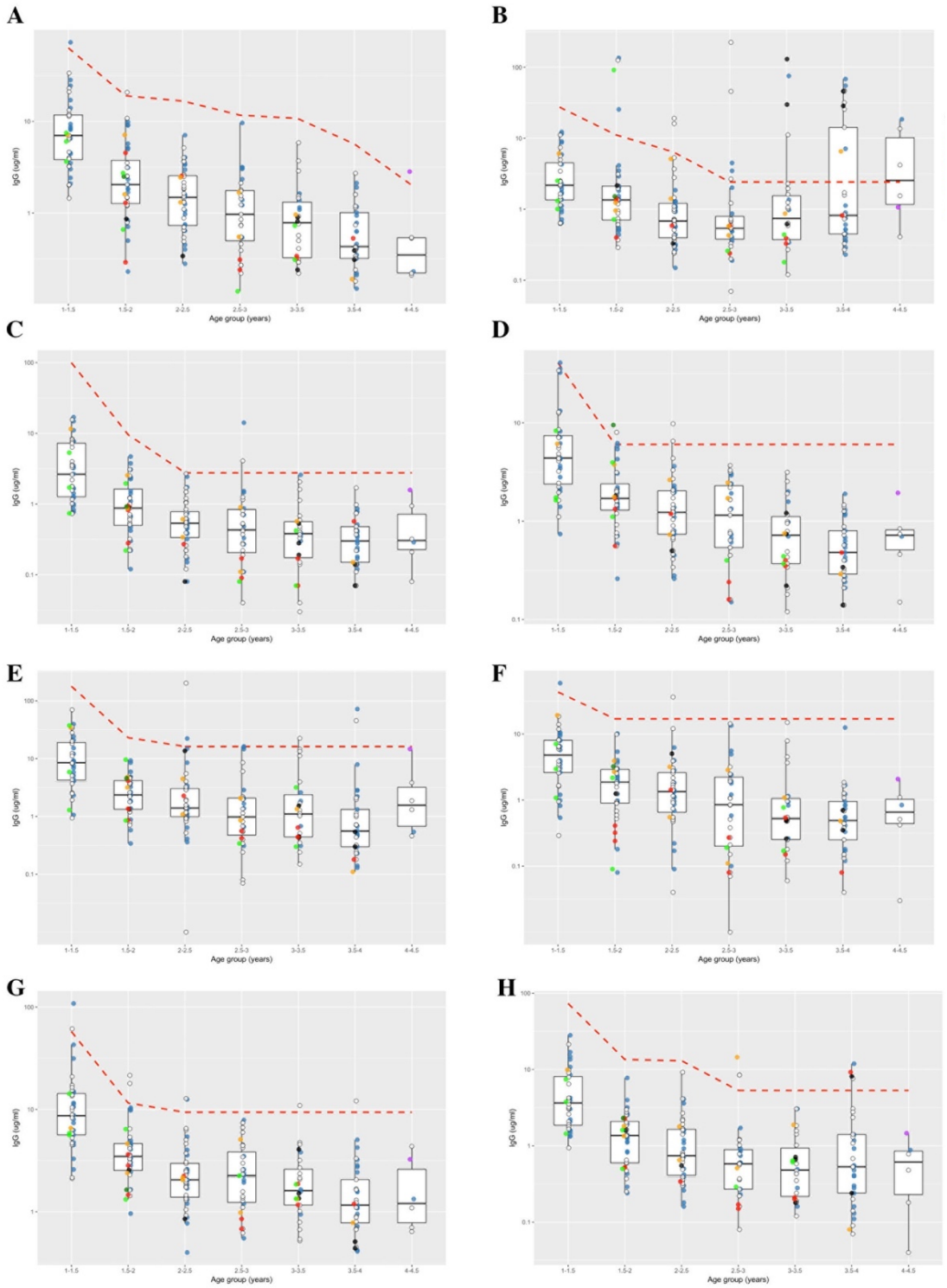


Figure 40. ***Detection of natural boosting events to serotypes 1 (A), 3 (B), 4 (C), 5 (D), 6A (E), 6B (F), 7F (G), and 9V (H) by sero-prevalence.***

Serum from PCV13 vaccinated UK children was analysed for IgG against serotype-specific pneumococcal capsular polysaccharide and concentrations plotted against age of sampling. Each point is coloured according to whether the child was also carrying serotypes 3 (black), 19A (light green), 19F (dark green), 6C (purple), 23A (orange), 23B (red), non-carrier (white) or NVT (blue). At the same time as serum collection a nasopharyngeal swab was collected and any detected pneumococci serotyped by DNA microarray. Natural boosting events were defined as those participants who had an antibody level above the threshold defined by the formula: 1.5 x (IQR+Q3). This threshold was applied across the age-quartiles until there was an increase in the threshold, at which point the threshold was kept at the lowest limit it reached before it increased.

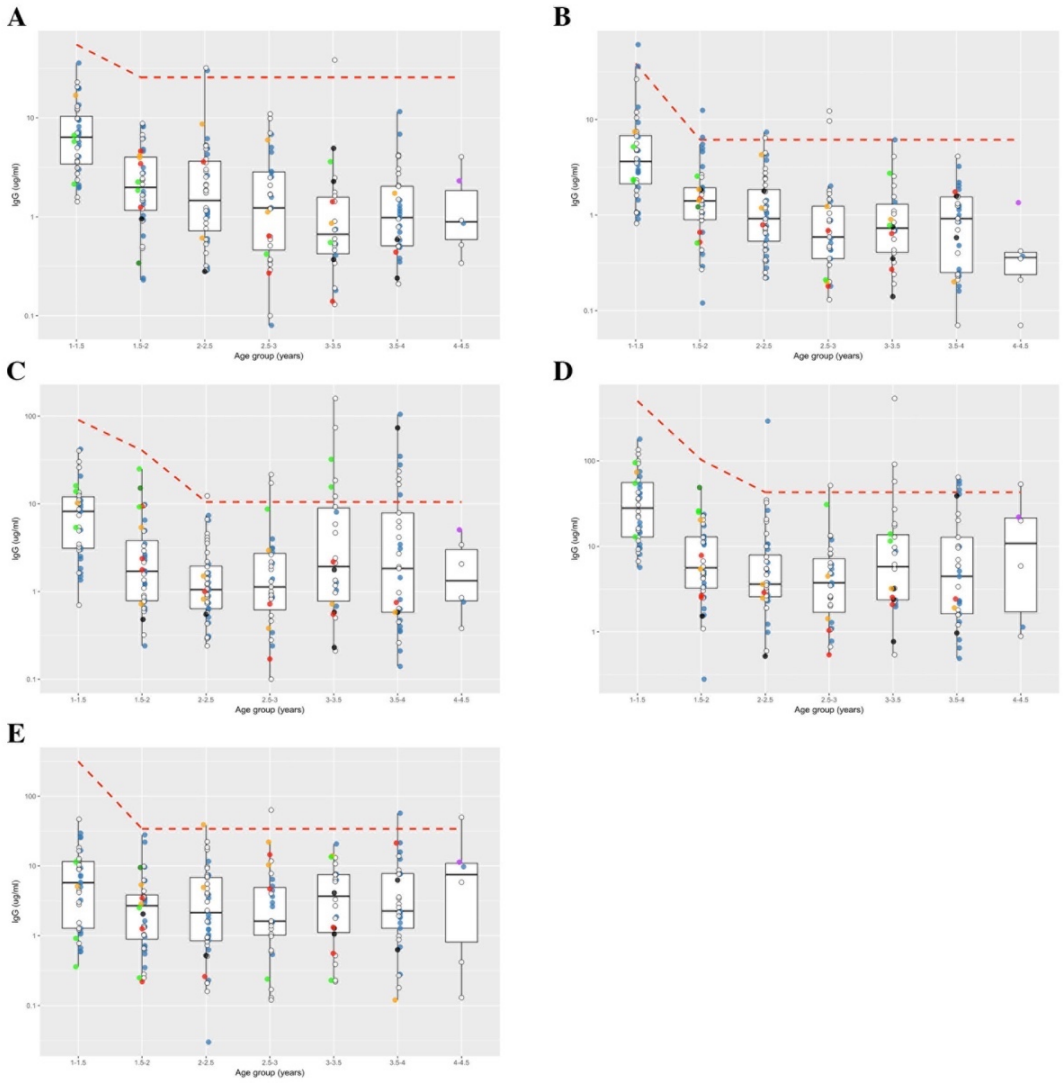


Figure 41. ***Detection of natural boosting events to serotypes 14 (A), 18C (B), 19A (C), 19F (D), and 23F (E) by sero-prevalence.***

Serum from PCV13 vaccinated UK children was analysed for IgG against serotype-specific pneumococcal capsular polysaccharide and concentrations plotted against age of sampling. Each point is coloured according to whether the child was also carrying serotypes 3 (black), 19A (light green), 19F (dark green), 6C (purple), 23A (orange), 23B (red), non-carrier (white) or NVT (blue). At the same time as serum collection a nasopharyngeal swab was collected and any detected pneumococci serotyped by DNA microarray. Natural boosting events were defined as those participants who had an antibody level above the threshold defined by the formula: 1.5 x (IQR+Q3). This threshold was applied across the age-quartiles until there was an increase in the threshold, at which point the threshold was kept at the lowest limit it reached before it increased.

Appendix C

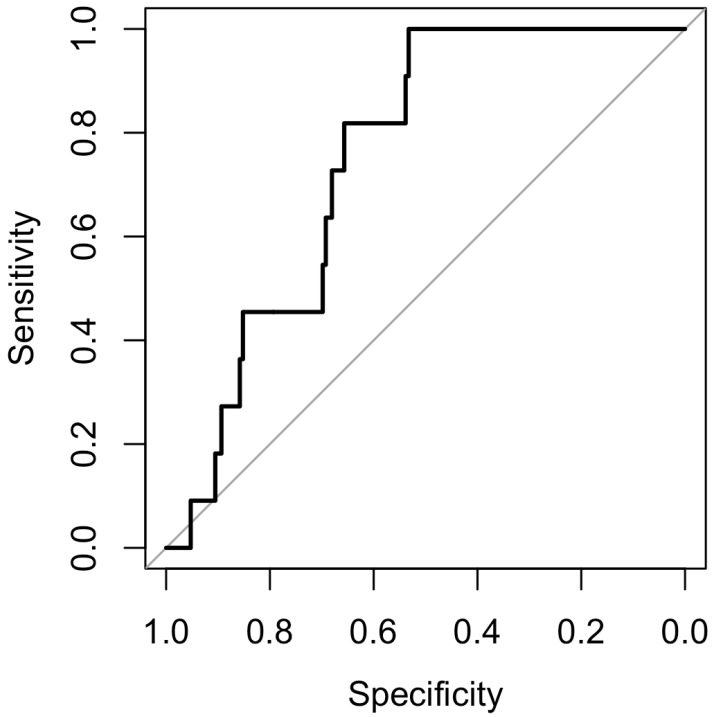


Figure 42. ***ROC curve demonstrating the performance of lytA copy number for differentiating Nepalese children admitted to Patan Hospital with pneumococcal pneumonia (defined as nasopharyngeal carriage of serotypes 1 or 5, a blood neutrophil count >7.5x10^9, and abnormal chest x-ray, or an abnormal chest x-ray and pneumococcus detected by blood culture) from those with who did not meet the criteria but had abnormal chest x-rays.***

AUC=0.75. A cut-off threshold of 486,416 *lytA* gene copies/ml confers a sensitivity of 64% and a specificity of 68%.

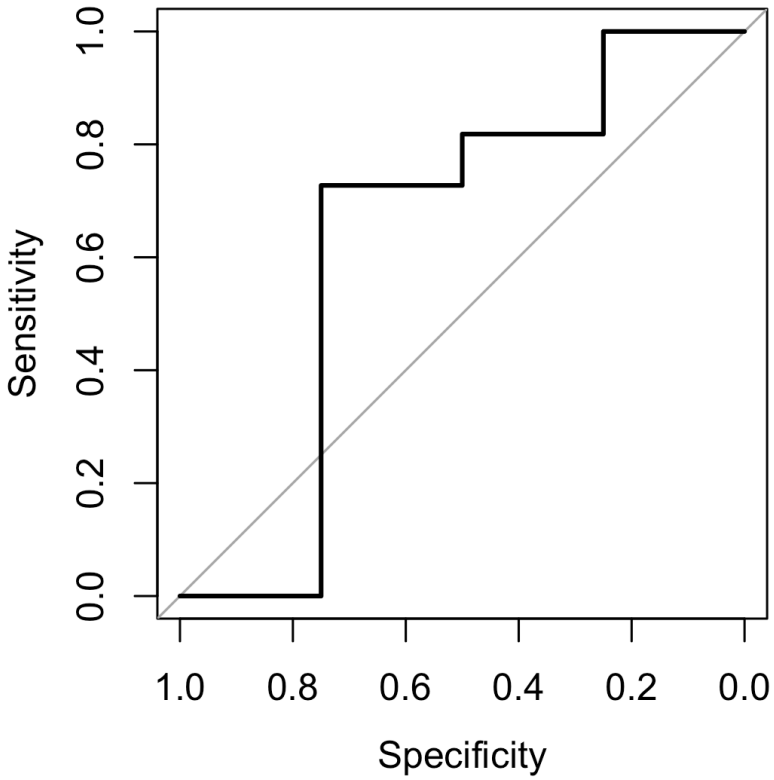


Figure 43. ***ROC curve demonstrating the performance of lytA copy number for differentiating Nepalese children admitted to Patan Hospital with pneumococcal pneumonia from those with normal chest x-rays that were carrying either serotypes 1, 5, or 14.***
AUC= 0.64. Using a cut-off of 3,048,551 *lytA* gene copies/ml confers a sensitivity of 73% and specificity of 75%.