Longitudinal colonisation by *Streptococcus pneumoniae* and nasopharyngeal microbial interactions in health and disease: A South African birth cohort study

Felix Sizwe Dube

Thesis presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Division of Medical Microbiology

UNIVERSITY OF CAPE TOWN

27 December 2015

DECLARATION

I hereby declare that the work on which the thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work or any part thereof is being, has been, or is to be submitted for another degree in this or any other University. The research described here was carried out in the Division of Medical Microbiology, Faculty of Health Sciences, University of Cape Town under the supervision of Professor Mark Nicol and Dr Lourens Robberts.

Felix Sizwe Dube

27 December 2015

ABSTRACT

Introduction: *Streptococcus pneumoniae* (the pneumococcus) is the most common cause of childhood pneumonia. Nasopharyngeal (NP) colonization by the pneumococcus is a necessary first step in the pathogenesis of pneumonia and yet the dynamic nature of pneumococcal colonization remains incompletely understood. In children, asymptomatic colonization of the nasopharynx by the pneumococcus is common and also serves as a reservoir for person-to-person transmission. We aimed to investigate in detail, the dynamics of pneumococcal nasopharyngeal carriage over the first year of life, in a cohort of South African children, particularly after implementation of the 13-valent pneumococcal conjugate vaccine (PCV-13). The study will further elucidate the interaction of *S. pneumoniae* with other respiratory pathogens and how such interactions may contribute development of severe disease.

Methods: (i) We first compared three NP swab-types; nylon flocked, polyethylene terephthalate (PET) and rayon for the optimal recovery of *S. pneumoniae*. (ii) In addition, a real-time multiplex PCR (rmPCR) was compared to a sequetyping method in order to assess their suitability for high throughput pneumococcal capsular typing. Whole genome re-sequencing was used to resolve discordant molecular serotyping and reference-standard Quellung results. (iii) Following optimization of specimen collection and processing protocols, NP swabs were prospectively collected from 137 infants enrolled in a birth cohort study during scheduled 2-weekly clinic visits and during each episode of pneumonia from May 2012 through April 2014. Paired NP swabs were either placed into skimmed milk tryptone glucose glycerol transport media (STGG) for culture-based detection methods or placed into PrimeStore™ for molecular based detection

methods. Following automated total nucleic acid extraction, a quantitative real-time PCR (qPCR) assay for the detection of the *S. pneumoniae* autolysin-encoding (*lyt*A) gene was performed and its specificity assessed by screening a pool of heterogeneous bacterial species. The sequetyping assay was used to type all the cultured *S. pneumoniae* and Quellung used for confirmation. (iv) A real-time multiplex reverse transcriptase PCR (FTDresp33) was employed to determine patterns of colonization/infection with other pathogens and hence to describe the interaction of S. pneumoniae with other members of the NP microbiome in a pilot study of 214 children presenting for care with symptoms suggestive of tuberculosis.

Results: (i) The recovery of *S. pneumoniae* from nylon flocked swabs was greater compared to both PET and rayon swabs when testing mock specimens. Similarly, higher bacterial loads were detected by qPCR from flocked swabs compared with PET swabs from healthy children. (ii) The rmPCR assay performed well for the 21 serotypes/-groups included in the assay and for serotype mixtures. However, in our study setting, a large proportion of serotypes were not detected by rmPCR. The sequetyping assay performed well, with specific exceptions, and may be more useful for regions where vaccine serotypes are less common. Whole genome re-sequencing identified structural differences in the *wzh* gene of some strains that were serotyped as 16F by Quellung, but incorrectly sequetyped as 9V. (iii) *S. pneumoniae* was identified in 54% (1809/3331) of the NP swabs collected from 137 intensively sampled infants during their first year of life. Of these, 90% (1637/1809) were typeable while 9%(172/1809) were non-typeable. The point prevalence of NP carriage at 2 weeks of age was 2%, reached a maximum prevalence of 60% at 24 weeks of age, and thereafter plateaued throughout the first year of life. At each sampling point, non-vaccine-types (NVT) accounted for

more than 80% of S. pneumoniae detected. The median time to first *S. pneumoniae* acquisition was 64 days (95% CI 56 – 91 days), with no difference observed between vaccine-type (VT) and NVT (p = 0.69). Serotype-specific acquisition rates ranged from 0.06 to 0.15 episodes / child days, whereas the carriage duration for all 48 different serotypes detected in this study was 1.2 to 29 days. Prior VT pneumococcal colonization before PCV-13 vaccination had no clear effect on the time to VT and NVT colonization post PCV-13 vaccination, p = 0.51 and p = 0.12 respectively. In contrast, NVT colonization prior to vaccination increased the hazard of acquiring NVT pneumococci after the first dose of PCV-13 by 3.6-fold (95% CI 1.95 - 6.52, p < 0.001). Interestingly, 'residual' carriage of VT serotypes did not appear to be affected by sequential doses of PCV-13. Children with older siblings were at increased risk of acquiring both VT and NVT pneumococci (HR = 1.83 [95% CI =1.66 1.00 - 2.73, p=0.04] and HR=1.27 - 2.64, p=0.001] respectively). (iv) Overall, at least one potential respiratory pathogen was detected in 97% of the respiratory specimens collected from the children with symptoms suggestive of tuberculosis. Besides S. pneumoniae (42%), the other common bacterial species detected were *Moraxella cartarrhalis, Haemophilus influenzae* and *Staphylococcus aureus* (64%, 29% and 22% respectively). Clinically relevant bacteria included *Mycoplasma pneumoniae, Bordetella pertussis* and *Chlamydia pneumoniae*. The most common viruses included Metapneumovirus (hMPV), Rhinovirus, Influenza C virus, Adenovirus, Cytomegalovirus and Coronavirus OC43. Whilst there was no clear separation between the pathogens present in the airways of children with and without PTB, C. pneumoniae, hMPV, coronavirus 043, influenza C virus, rhinovirus and cytomegalovirus were associated with PTB. On the other hand, Pneumocystis jirovecii, H. influenzae spp, Respiratory syncytial virus, M. pneumoniae, influenza B virus and enteroviruses were more common amongst children without TB.

Conclusion: We describe, in detail, the dynamic nature of NP carriage in relation to vaccination. We show that "residual" carriage of VT strains was driven by HIV-exposure (not infection) and that prior colonization by NVT pneumococcus substantially increased the risk of subsequent pneumococcal acquisition.

ACKNOWLEDGEMENT

I wish to acknowledge the unfailing support and guidance of my supervisors Prof. Mark Nicol[1] and Dr. Lourens Robberts[1]. Your invaluable insights and patience have truly made my PhD experience rewarding, enjoyable and worthwhile. Mark: Thank you for re-igniting my passion for microbiology, the numerous opportunities, encouragement and more experiments even when there wasn't time left to perform them. Lourens: Thank you for being a great listener and valuable insights that will go a long way in shaping my character.

I am exceedingly grateful to the generous support from the South African National Research Foundation as well as additional support from the Bill and Melinda Gates Foundation through the Division of Medical Microbiology. I am equally indebted to the support from The Wellcome Trust, Great Britain for the financial support to attend the Advanced Courses in Molecular Approaches to Clinical Microbiology (10-16 September 2011 Blantyre, Malawi) and Pathogen genomes workshop (10-16 September, 2014 Blantyre, Malawi).

I further wish to thank the following individuals for their technical support and expertise: Dr. Suzan van Mens[2] for the technical support in the pneumococcal capsular typing tools validation and her unwavering friendship (….and for naming one of her children after me). Dr. Nicole Wolter[3] for the intellectual exchange and supervision with

[1]Division of Medical Microbiology, Faculty of Health Sciences, University of Cape Town, Cape
[2]Department of Medical Microbiology and Immunology, St Antonius Hospital, Nieuwegein, the Netherlands.
[3]Centre for Respiratory Diseases and Meningitis (CRDM), National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa, & School of Pathology, Faculty of Health Sciences, University of the Witswatersrand, Johannesburg, South

the pneumococcal Quellung testing and confirmatory PCR testing through the Centre for Respiratory Diseases and Meningitis (CRDM), National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, Johannesburg, South Africa. Dr Anne von Gottberg[3] for the oversight and preliminary phenotypic characterization of pneumococcal strains provided through the NICD. Associate Prof. Sugnet Lubbe[4], Mr. Polite Nduru[5] and Mr. Jordache Ramjith[5] for the statistical analysis. Dr. Mushal Ali[3] and Dr. Joseph Mafofo[6] for the training and bioinformatics support. Dr. Paul Nicol[7] for the help in bioinformatics and writing the script for the Sequetyper and FTD Analyser. Dr. Mamadou Kaba[1] for the timely critical comments.

I am highly indebted to all the clinical research staff involved in the Drakenstein Child Lung Health Study for collection of samples and the children and parents for participating in the study. I further wish to thank all the staff, affiliates, friends and colleagues, both past and present, from the Division of Medical Microbiology for the assistance, experience and friendship over the years. In particular in wish to thank (in no particular order), Ms. Sleendile Mbhele (your laughter is a constant reminder of immeasurable joy), Dr. Veronica Allen, Ms. Menna Duyver (my adopted sister, thanks for being a gentle soul and ever so present a friend), Ms. Shantelle Classen, Ms. Shima Abdugalder, Dr. Adebayo Shittu (my Olga), Ms. Malefu Moleki and Mr. Babu Muhammed, Ms. Samantha Africa, Mr. Jermaine Khumalo (for the silent understanding and always being their for each other), Mr. Rendani Manenzhe, Ms. Sibongile Zaze (MHSRIP). You

---

Africa
[4]Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa
[5]Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, University of Cape Town, South Africa
[6]Centre for Proteomic and Genomic Research (CPGR), Cape Town, South Africa.
[7]The State Agricultural Biotechnology Centre, Murdoch, University, Murdoch, Australia,

made my PhD experience a pleasant and exciting experience. Forever indebted to you all.

On a personal note, I would like to honor the memory of my late mom and dad to whom I dedicate this piece. To my wife Nchimunya and daughters Sinikiwe and Unathi, you guys have been a pillar and source of motivation through it all and thank you for always being so patient with me and sacrificing your dreams. It is your love, encouragement, prayers, support, motivation and cherished moments spent together that sustained me. I would also like to thank my Creator for His unwavering love, mercy, strength, wisdom and provision throughout my academic pursuits. Thank you for shaping and moulding me!

PEER-REVIEWED PUBLICATIONS FROM THIS STUDY

- **Dube FS**, Kaba M, Whittaker E, Zar HJ, Nicol MP. Detection of *Streptococcus pneumoniae* from Different Types of Nasopharyngeal Swabs in Children. PLoS One.2013 Jun 26;8(6): e68097. Print 2013. PubMed PMID: 23840817; PubMed Central PMCID: PMC3694050.

- **Felix Dube** and Sugnet Gardner-Lubbe: Visualisation of quadratic discriminant analysis and its application in exploration of microbial interactions. BioData Mining 2015, 8:8 doi:10.1186/s13040-015-0041-9

- **F S Dube**, S P Van Mens, Lourens Robberts, N Wolter, J Mafofo, P Nicol, H Zar, M P Nicol Comparison of a Multiplex Real-Time PCR and Sequetyping Assay for Pneumococcal Serotyping PLoS One. 2015 Sep 3; 10(9): e0137349. Doi: 10.1371/journal.pone.0137349

- **Felix Dube**, Mamadou Kaba, Lourens Robberts, Lemese Ah Tow, Sugnet Lubbe, Heather J. Zar, Mark P. Nicol: Respiratory Microbes In Children Hospitalised With Suspected Pulmonary Tuberculosis In Cape Town, South Africa. Submitted BMC ID March 2016

PEER-REVIEWED CONFERENCE PROCEEDINGS

- **Felix S. Dube,** Khomotso Mmakola, Jordache Ramjith, Polite Nduru, F.J. Lourens Robberts, Heather J. Zar, Mark P. Nicol: Dynamics of serotype-specific nasopharyngeal carriage of *Streptococcus pneumoniae* in the period preceding pneumonia: the Drakenstein Child Health Study. Abstract Accepted: 10th International Symposium On Pneumococci & Pneumococcal Diseases, 26th – 30th June, 2016| Glasgow, Scotland

- **Felix S. Dube**, Jodarch Ramjith, Sugnet Lubbe, Nicole Wolter, Heather J. Zar, Mark P. Nicol: Longitudinal Nasopharyngeal Colonisation With *Streptococcus Pneumoniae* In The First Year Of Life: An Intensively Sampled Pcv-13-Vaccinated Birth Cohort. Abstract Accepted: 10th International Symposium On Pneumococci & Pneumococcal Diseases, 26th – 30th June, 2016| Glasgow, Scotland

- **Felix S. Dube**, Jodarch Ramjith, Lourens Robberts, Sugnet Lubbe, Suzan P. van Mens, Nicole Wolter, Heather J. Zar, Mark P. Nicol: Impact of PCV13 on the rate of *Streptococcus pneumoniae* acquisition and duration of carriage in a South African birth cohort. Abstract Oral presentation at the 6th Federation of Infectious Diseases Society of South Africa, 5 – 8th October 2015, Durban, South Africa.

- **Felix S. Dube**, Anat Melamed, Nicolas A. Gillet, Nonhlanhla Lunjani, Daniel J. Laydon, Charles R.M. Bangham, Carol Hlela: Human T-Lymphotropic Virus-1 Clonality In Infective Dermatitis associated with HTLV-1 (IDH) And After Spontaneous

Resolution. Abstract poster presentation at the 6th Federation of Infectious Diseases Society of South Africa, 5 – 8th October 2015, Durban, South Africa.

- **Felix S. Dube**, Lourens Robberts, Suzan P. van Mens, Nicole Wolter, Heather J. Zar, Mark P. Nicol. *Streptococcus pneumoniae* nasopharyngeal colonization, serotype distribution and impact of vaccination during the first year of life: A South African birth cohort study. Presented as a poster at the University of Cape Town's Faculty of Healthy Sciences Research Day, 3rd September 2015.

- **Felix Dube** and Sugnet Gardner-Lubbe: Functional Data Analysis for the investigation of longitudinal response patterns in health care. Oral presantion at the International Federation of Classification Societies (IFCS) conference 6 - 8 July 2015, in Bologna, Italy.

- **Felix S. Dube**, Lourens Robberts, Suzan P. van Mens, Nicole Wolter, Heather J. Zar, Mark P. Nicol. *Streptococcus pneumoniae* nasopharyngeal colonization, serotype distribution and impact of vaccination during the first year of life: A South African birth cohort study. Young Investigator Oral Presentation, American Society of Microbiology 2015 (ASM2015) general meeting, New Orleans, Louisiana, May 30-June 2, 2015

- S P Van Mens, **F S Dube**, H Zar, M P Nicol: Comparison of multiplex real-time PCR and sequetyping for pneumococcal serotyping in South African, a low to middle income setting with high PCV13 coverage. International Society for Infectious Diseases (ICID), Cape Town, South Africa; 04/2014

- **Felix Dube**, Mamadou Kaba, Lourens Robberts, Lemese Ah Tow, Samantha Africa, Sugnet Lubbe, Heather J. Zar, Mark P. Nicol: Microbiology of the nasopharynx in children hospitalized with suspected pulmonary tuberculosis. International Society for Infectious Diseases (ICID), Cape Town, South Africa; 04/2014

- **Dube FS**, Kaba M, Whittaker E, Zar HJ, Nicol MP. Detection of *Streptococcus pneumoniae* from Different Types of Nasopharyngeal Swabs in Children. 5th Federation of Infectious Diseases Society of Southern Africa (FIDSSA), 10 – 12 October 2013, Durban, South Africa.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A_{600}	Absorbance at 600nm
6-FAM	6 - Carboxyfluorescein
ALRI	Acute Lower Respiratory Tract Infection
ALRI	Acute Lower Respiratory Tract Infections
ALRTI	Acute Lower Respiratory Tract Infections
Amp	Ampicillin
ANOVA	Analysis of variance
AOM	Acute otitis media
β	Beta
BA	Blood agar
BHI	Brain Heart infusion
BLAST	Basic Local Alignment Search Tool
Bp	Base pair(s)
CBP	Choline binding protein
CI	Confidence Interval
CMV	Cytomegalovirus
CFU	Colony forming units
COPD	Chronic Obstructive Pulmonary Disorder
CLSI	Clinical Laboratory Standard Institute
CSF	cerebrospinal fluid
CPS	Capsule
Colgent	Columbia gentamycin
*cps*B	Capsular biosynthetic gene cluster B

DNA	deoxyribonucleic acid
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylene diamine tetraacetic acid
EPI	Expanded Programme On Immunization
EtBr	Ethidium bromide
Ery	Erythromymicin
g	Gravitational units
h	hours
HiB	Haemophilus Influenzae Type B
HIV	Human Immunodeficiency Virus
IPD	Invasive pneumococcal disease
IPTG	Isopropyl-β-D-Thiogalactopyranoside
Kb	Kilobase pair (s)
LB	Luria Bertani broth
LPS	Lipopolysaccharide
LPxTG	Leucine-proline-x-threonine-glycin binding motif
LRT	Lower Respiratory Tract
LRTI	Lower Respiratory Tract Infections
lytA	Autolysin A
$MgCl_2$	Magnesium chloride
min	Minute(s)
MLST	Multilocus sequence type
Nan (A, B, C)	Neuraminidase (A, B, C)
NCBI	National Center for Biotechnology Information at the U.S National Library of Medicine, Bethesda, Maryland, USA

nt	Nucleotide(s)
N-terminus	Amino terminus
OD	Optical Density
OM	Otitis media
O/N	Overnight
ORF	Open reading frame
PBS	Phosphate buffered saline
Para (1,2,3)	Human Parainfluenza Virus (1,2,3)
PCR	Polymerase chain reaction
PCV	Pneumococcal Polysaccharide-Diptheria CRM_{197} Protein Conjugate Vaccine
PPS	Pneumococcal Polysaccharide Vaccine
PCA	principal component analysis
Ply	Pneumolysin
PspA	Pneumococcal surface protein A
PspC	Pneumococcal surface protein C
RNA	Ribonucleic acid
RT	Room temperature
RT-PCR	Reverse transcriptase polymerase chain reaction
rpm	Revolution Per Minute
rRNA	Ribosomal Ribonucleic Acid
RSV	Respiratory Syncytial Virus
SNP	Single nucleotide polymorphism
ST	Sequence type
STGG	Skim Milk-Tryptone-Glucose-Glycerol

TAE	Tris-Acetate-EDTA
TH	Todd-Hewitt broth
THY	Todd-Hewitt broth supplemented with yeast extract
URTI	Upper Respiratory Tract Infections
VT	Vaccine serotypes
WHO	World health organisation
UV	Ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
α	alpha
µg	Microgram
µL	Microlitre

CHAPTER 1

General Introduction

1.1 Brief History

Streptococcus pneumoniae (the pneumococcus), is a gram-positive, alpha-hemolytic, facultative anaerobic bacterium that colonises the nasopharynx of many healthy children and adults, and causes disease in some of those colonised (1,2). George Miller Sternberg initially isolated the bacterium from the blood of a rabbit that he had inoculated with his own saliva in 1885 and named it *Microbe septicemique du salive* (3). Louise Pasteur independently isolated the bacterium after Sternberg and referred to the organism as *Micrococcus pasteur* (4). It was later in the 1920s that the organism was assigned the name *Diplococcus pneumoniae* due to its characteristic paired appearance on gram-stained sputum. Its current nomenclature, *Streptococcus pneumoniae,* is derived from the distinct chain-like (Streptococcal) morphology when grown in liquid media (3,4).

Since its discovery, research on the pneumococcus has led to many important contributions to the field of science and medicine. These discoveries include the Gram stain technique (5) and Quellung reaction; the latter remains the gold standard for pneumococcal serotyping (6,7). Experimental work on the pneumococcal capsular polysaccharide (CPS) forms the basis of current pneumococcal vaccine formulations (8). Frederick Griffith's experimental work on pneumococcal virulence determinants was key in demonstrating that pneumococci have the ability to take-up exogenous DNA from closely related species and undergo natural transformation (9). Oswald Avery and colleagues later showed that deoxyribonucleic acid (DNA) was the 'transforming principle' by demonstrating that treatment with DNase attenuated transformation *in vitro* (8). Avery et al further showed that the CPS was comprised of complex cross-

linked sugar molecules (8). The pneumococcus is highly recombinogenic; this genomic plasticity permits strains to adapt to antibiotic- and vaccine selective pressures.

1.2 Background and Rationale

The pneumococcus is a leading cause of mild upper respiratory tract infections such as otitis media and sinusitis, and severe disease such as pneumonia, sepsis and meningitis (10–12). Patients at risk include those at the extremes of age and the immunocompromised, particularly those affected by cell-mediated immune deficiencies (12–14). Pneumonia is a leading cause of death in children under the age of five years, accounting for approximately 17% of 6·3 million child deaths in 2010 (15). Childhood pneumonia further contributes to the public health burden, with 120 million episodes occurring in children under the age of five years; of these, 14 million progress to severe episodes. In essence, the highest proportion (81%) of pneumonia-related deaths occur in the first two years of life (15). According to the 2010 estimates, the incidence is particularly high in Africa where 36 million episodes result in more than 600,000 pneumonia-related deaths each year (10,16,17).

Despite high pneumococcal immunization coverage, the incidence of pneumonia (0.27 episodes per child-year, [e/cy]) and of severe pneumonia among South African children in a semi-rural setting remains disproportionately high, with almost a third of children developing pneumonia in the first year of life (16). The high incidence of pneumonia in our setting is more than double the estimated nationwide incidence for South African children younger than five years (0.14 e/cy) and higher than the median estimate in developing countries of 0.28 e/cy (Interquartile range, IQR, 0.21-0.71 e/cy), both

incidences were calculated prior to implementation of pneumococcal conjugate vaccine (11). However these differences may be due to differences in the rigor of case ascertainment and reporting. The observed incidence of pneumonia is substantially higher than the reported incidence (ranged from 0.006 – 0.085 e/cy) from in children from high-income countries (17). Moreover, the incidence of pneumonia is highest in children between 1 – 6 months (Incidence rate ratio, 3.00 [95% Confidence interval, CI, 1.41-7.67]) (16), suggesting increased vulnerability of younger children to pneumococcal disease.

Nasopharyngeal (NP) colonisation by the pneumococcus is a necessary first step in the development of both mucosal and invasive disease and yet the dynamic nature of pneumococcal colonisation remains incompletely understood (1,2,12). The pneumococcus shares this ecological niche with more than 700 other microbial species including resident flora, transient colonizers and pathogenic species (18). It is believed that intra- and interspecies competition within this ecological niche contributes to fluctuations in pneumococcal carriage over time (19,20). In children, asymptomatic colonisation by the pneumococcus is common, serving as a reservoir for person-to-person transmission in the community (21). A complex interplay between pneumococcal and host factors determine whether the colonising pneumococcus is cleared by the host's immune system or progress to invasive disease (2). The outcome of this interaction is further influenced by vaccination and other multifactorial risk factors (11,12,22,23).

1.3 Thesis Outline

CHAPTER 2: NP colonisation by the pneumococcus is key to understanding the pathogenesis of pneumococcal disease and is increasingly used as an endpoint for pneumococcal vaccine studies. It is important to establish the optimal strategy for pneumococcal recovery from NP specimens. Sampling of the nasopharynx may be achieved using an NP swab or aspirate. NP swabs are preferred as the procedure is simpler, quicker and better tolerated by children. The ideal swab would be highly absorbent, maintain the viability of microorganisms present, release maximum specimen material into culture broth or transport medium and not inhibit downstream applications such as culture and nucleic acid amplification. However there have not been systematic studies to determine the best available swab for such applications. This chapter therefore describes experiments comparing pneumococcal recovery from nylon flocked, polyethylene terephthalate (PET) and rayon swabs by quantitative culture and *lytA*-targeted real-time polymerase chain reaction (qPCR).

CHAPTER 3: The CPS is an important virulence factor as un-encapsulated strains are virtually absent from invasive pneumococcal disease (IPD) (9,24,25). The CPS protects the pneumococcus from ingestion by host phagocytic cells and promotes colonisation by limiting mucus-mediated clearance of the pneumococci (3,24–26). CPS further facilitates person-to-person transmission by protecting the pneumococcus from desiccation and mechanical damage, enabling the bacterium to survive during transmission in an aerosolised droplet (27). There are over 98 different pneumococcal serotypes differentiated by the antigenic CPS (28). Individual serotypes are associated with differing potential to cause invasive disease, antimicrobial drug-resistance and

have different global distribution patterns (9,22,25,26,29). Genetic shifts in the population structure of pneumococci can result from serotype replacement, expansion of pre-existing clones, acquisition of novel clones, and from capsular switching - a change of the pneumococcal capsule by a clone due to alteration or exchange of its capsular biosynthetic gene cluster (*cps*) (30).

More practical, higher throughput serotyping techniques are required for expanding public health laboratory services in many areas of the world to support growing disease control programs and epidemiological surveillance. Conventional serotyping methods are costly, time-consuming and error prone. Current DNA-based methods require large number of serotype-specific primer pairs, while microarrays are costly and only available in selected laboratories. Chapter 3 therefore presents data from head-to-head comparisons of the accuracy of two methods suitable for large-scale serotyping, viz. sequetyping and multiplex real-time PCR. The sequetyping assay utilizes a single-primer PCR and sequence-based method targeting the capsular regulatory gene *wzh,* and can correctly identify a high number of serotypes, including those included in the 13-valent conjugate vaccine (31). The multiplex real-time PCR targets 21 different serotypes/-groups including all PCV-13 serotypes as well as additional serotypes/-groups (32).

CHAPTER 4 & 5: A better understanding of the effects of serotype-specific vaccination on the dynamics of pneumococcal serotype-specific colonisation is needed for monitoring vaccine efficacy and potential serotype replacement. A 13-valent pneumococcal conjugate vaccine (PCV-13, Prevenar®, Wyeth Pharmaceuticals Inc.) was introduced in South Africa in June 2011 to replace the 7-valent pneumococcal conjugate vaccine (PCV-7, Prevenar®, Wyeth Pharmaceuticals Inc.), which was introduced in April

2009 (33). PCV-13 is administered in a 2+1-dosing schedule at 6, 14 and 36 weeks of age (33). This immunisation schedule was initially based on recommendations from the ministerial-appointed National Advisory Group for Immunisation (NAGI) in South Africa (34) and subsequent evidence of its effectiveness in HIV-exposed but uninfected South African children (35).

PCVs are effective in preventing serotype-specific invasive disease; however, there is growing concern over replacement of vaccine serotypes with non-vaccine serotypes in carriage and invasive clinical isolates (20,36–39). Our understanding of the serotype-specific mucosal immunity induced by PCV is incomplete. The effect of sequential vaccination on longitudinal colonisation with vaccine vs. non-vaccine serotypes in individual children is largely unknown. Furthermore, the effect of serotype-specific colonisation (i.e., the pneumococcus acting as a natural immunizing agent) on subsequent acquisition of the same or different serotypes and on vaccine effectiveness remains poorly understood. Chapter 4 describes longitudinal pneumococcal colonisation patterns and serotype distribution during the first year of life among South African infants participating in a birth cohort (Drakenstein Child Health Study). In addition, this chapter details serotype-specific acquisition rates and duration of pneumococcal carriage. Chapter 5 specifically investigates risk factors for acquisition of pneumococcal carriage including the short-term impact of sequential PCV-13 immunisation on carriage.

CHAPTER 6: The microbial composition of the NP microbiome and interactions within this niche may play a key role in determining the outcome of infection in children. Recent studies into the pathobiome have highlighted the impediments to the long-established concept of "one microbe - one disease", as embodied in Koch's postulates.

These studies report an increase in the detection of polymicrobial infections in children with respiratory tract infections, this includes children with suspected pulmonary TB and those infected with human immunodeficiency virus (HIV) (40–47). The diagnosis of TB in children is largely made on the basis of clinical and radiological presentations, which may be non-specific (48). In children, tuberculosis (TB) may present as either an acute or chronic lower respiratory tract infection, or with extrapulmonary manifestations (49). *Mycobacterium tuberculosis* is increasingly recognized as a primary cause or underlying comorbidity amongst children with severe pneumonia, particularly in TB and HIV endemic settings (50).

Little is known about the range of respiratory pathogens that may be detected in children presenting with suspected TB, either as the primary cause of respiratory symptoms in children who do not have TB, or as co-pathogens in children who do have TB. Although interactions between specific bacterial and viral pathogens occupying the NP microbiome have been described, there are no published data providing a comprehensive picture of the bacterial and viral composition of the NP microbiome in children with TB. This chapter therefore describes the identification of potential respiratory tract pathogens (other than *M. tuberculosis*) from NP samples collected from in a cohort of children presenting at Red Cross Children War Memorial Hospital with suspected TB. A multiplex real-time reverse transcriptase PCR (Fast-Track Diagnostics, Luxembourg) assay was applied to detect (and quantify) up to 33 potential respiratory pathogens including 8 bacterial species, 11 viruses and a fungus. This work was further extended to developing and validating a novel statistical technique (quadratic discriminant analysis) to understanding polymicrobial pathogen interactions.

CHAPTER 7: This section will summarize the key findings from these studies in the context of existing knowledge and highlight how this work has advanced knowledge in terms of NP sample collection, capsular serotyping, determinants of pneumococcal carriage and the role of NP microbial interactions in health and disease. This chapter will also highlight our approach to future studies, which will include increasing the sample size of the longitudinal cohort, application of whole genome sequencing to further study the molecular epidemiology of the pneumococcus in the cohort and assessment of the pneumococcal antimicrobial resistome using an enrichment-based next generation sequencing approach.

1.4 REFERENCES

1. Gray BM, Turner ME, Dillon HC (1982) Epidemiologic studies of *Streptococcus pneumoniae* in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. Am J Epidemiol 116: 692–703.
2. Simell B, Auranen K, Käyhty H, Goldblatt D, Dagan R, et al. (2012) The fundamental link between pneumococcal carriage and disease. Expert Rev Vaccines 11: 841–855.
3. Watson DA, Musher DM, Jacobson JW, Verhoef J (1993) A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. Clin Infect Dis 17: 913–924.
4. Austrian R Pneumococcus: the first one hundred years. Rev Infect Dis 3: 183–189.
5. Gram H (1884) Uber die isolierte Farbung der Schizomyceten in Schnitt- und Trockenpraparaten. Fortschr Med 2.
6. Satzke C, Dunne EM, Porter BD, Klugman KP, Mulholland EK (2015) 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; discussion e1001903.
7. Satzke C, Turner P, Virolainen-Julkunen A, Adrian P V., Antonio M, et al. (2013) Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. Vaccine 32: 165–179.
8. Avery OT, Macleod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. J Exp Med 79: 137–158.
9. Griffith F (1928) The Significance of Pneumococcal Types. J Hyg (Lond) 27: 113–159.
10. Wang H, Liddell C a., Coates MM, Mooney MD, Levitz CE, et al. (2014) Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet.
11. Rudan I, O'Brien KL, Nair H, Liu L, Theodoratou E, et al. (2013) Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. J Glob Health 3: 010401..
12. Bogaert D, De Groot R, Hermans PWM (2004) *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis 4: 144–154.
13. Coles CL, Sherchand JB, Khatry SK, Katz J, Leclerq SC, et al. (2009) Nasopharyngeal carriage of *S. pneumoniae* among young children in rural Nepal. Trop Med Int Health 14: 1025–1033. doi:10.1111/j.1365-3156.2009.02331.x.
14. Antonio M, Dada-Adegbola H, Biney E, Awine T, O'Callaghan J, et al. (2008) Molecular epidemiology of pneumococci obtained from Gambian children aged 2-29 months with invasive pneumococcal disease during a trial of a 9-valent pneumococcal conjugate vaccine. BMC Infect Dis 8: 81.
15. Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, et al. (2013) Global burden of childhood pneumonia and diarrhoea. Lancet 381: 1405–1416.
16. le Roux DM, Myer L, Nicol MP, Zar HJ (2015) Incidence and severity of childhood

pneumonia in the first year of life in a South African birth cohort: the Drakenstein Child Health Study. Lancet Glob Heal 3: e95–e103.
17. Madhi SA, De Wals P, Grijalva CG, Grimwood K, Grossman R, et al. (2013) The burden of childhood pneumonia in the developed world: a review of the literature. Pediatr Infect Dis J 32: e119–e127.
18. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE (2005) Defining the normal bacterial flora of the oral cavity. J Clin Microbiol 43: 5721–5732.
19. Auranen K, Mehtälä J, Tanskanen A, S Kaltoft M (2010) Between-strain competition in acquisition and clearance of pneumococcal carriage--epidemiologic evidence from a longitudinal study of day-care children. Am J Epidemiol 171: 169–176.
20. Weinberger DM, Malley R, Lipsitch M (2011) Serotype replacement in disease after pneumococcal vaccination. Lancet 378: 1962–1973.
21. Abdullahi O, Nyiro J, Lewa P, Slack M, Scott JAG (2008) The descriptive epidemiology of *Streptococcus pneumoniae* and Haemophilus influenzae nasopharyngeal carriage in children and adults in Kilifi district, Kenya. Pediatr Infect Dis J 27: 59–64.
22. Sjöström K, Spindler C, Ortqvist A, Kalin M, Sandgren A, et al. (2006) Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. Clin Infect Dis 42: 451–459.
23. Walaza S, Tempia S, Dawood H, Variava E, Moyes J, et al. (2015) Influenza virus infection is associated with increased risk of death amongst patients hospitalized with confirmed pulmonary tuberculosis in South Africa, 2010¿2011. BMC Infect Dis 15: 26.
24. Jansen AGSC, Rodenburg GD, van der Ende A, van Alphen L, Veenhoven RH, et al. (2009) Invasive pneumococcal disease among adults: associations among serotypes, disease characteristics, and outcome. Clin Infect Dis 49: e23–e29.
25. Browall S, Norman M, Tångrot J, Galanis I, Sjöström K, et al. (2014) Intraclonal variations among *Streptococcus pneumoniae* isolates influence the likelihood of invasive disease in children. J Infect Dis 209: 377–388.
26. Morona JK, Miller DC, Morona R, Paton JC (2004) The effect that mutations in the conserved capsular polysaccharide biosynthesis genes cpsA, cpsB, and cpsD have on virulence of *Streptococcus pneumoniae*. J Infect Dis 189: 1905–1913.
27. Walsh RL, Camilli A (2011) *Streptococcus pneumoniae* Is Desiccation Tolerant and Infectious upon Rehydration. MBio 2: e00092–11 – e00092–11.
28. Jauneikaite E, Tocheva AS, Jefferies JMC, Gladstone RA, Faust SN, et al. (2015) Current methods for capsular typing of *Streptococcus pneumoniae*. J Microbiol Methods 113: 41–49.
29. Weinberger DM, Trzciński K, Lu Y-J, Bogaert D, Brandes A, et al. (2009) Pneumococcal capsular polysaccharide structure predicts serotype prevalence. PLoS Pathog 5: e1000476.
30. Beall BW, Gertz RE, Hulkower RL, Whitney CG, Moore MR, et al. (2011) Shifting genetic structure of invasive serotype 19A pneumococci in the United States. J Infect Dis 203: 1360–1368.
31. Leung MH, Bryson K, Freystatter K, Pichon B, Edwards G, et al. (2012) Sequetyping: Serotyping *Streptococcus pneumoniae* by a Single PCR Sequencing Strategy. J Clin Microbiol 50: 2419–2427. doi:10.1128/JCM.06384-11.
32. Pimenta FC, Roundtree A, Soysal A, Bakir M, du Plessis M, et al. (2013) Sequential triplex real-time PCR assay for detecting 21 pneumococcal capsular serotypes

that account for a high global disease burden. J Clin Microbiol 51: 647–652.
33. Madhi SA, Cohen C, von Gottberg A (2012) Introduction of pneumococcal conjugate vaccine into the public immunization program in South Africa: translating research into policy. Vaccine 30 Suppl 3: C21–C27.
34. Schoub BD, Ngcobo NJ, Madhi S (2010) The National Advisory Group on Immunization (NAGI) of the Republic of South Africa. Vaccine 28 Suppl 1: A31–A34.
35. Cohen C, von Mollendorf C, de Gouveia L, Naidoo N, Meiring S, et al. (2014) Effectiveness of 7-Valent Pneumococcal Conjugate Vaccine Against Invasive Pneumococcal Disease in HIV-Infected and -Uninfected Children in South Africa: A Matched Case-Control Study. Clin Infect Dis 59: 808–818.
36. Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, et al. (2010) Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 201: 32–41.
37. Miller E, Andrews NJ, Waight PA, Slack MP, George RC (2011) Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis 11: 760–768.
38. Millar E V, O'Brien KL, Watt JP, Bronsdon MA, Dallas J, et al. (2006) Effect of community-wide conjugate pneumococcal vaccine use in infancy on nasopharyngeal carriage through 3 years of age: a cross-sectional study in a high-risk population. Clin Infect Dis 43: 8–15.
39. Huang SS, Hinrichsen VL, Stevenson AE, Rifas-Shiman SL, Kleinman K, et al. (2009) Continued impact of pneumococcal conjugate vaccine on carriage in young children. Pediatrics 124: e1–e11.
40. McNally LM, Jeena PM, Gajee K, Thula SA, Sturm AW, et al. (2007) Effect of age, polymicrobial disease, and maternal HIV status on treatment response and cause of severe pneumonia in South African children: a prospective descriptive study. Lancet 369: 1440–1451.
41. Honkinen M, Lahti E, Österback R, Ruuskanen O, Waris M (2012) Viruses and bacteria in sputum samples of children with community-acquired pneumonia. Clin Microbiol Infect 18: 300–307.
42. Hamano-Hasegawa K, Morozumi M, Nakayama E, Chiba N, Murayama SY, et al. (2008) Comprehensive detection of causative pathogens using real-time PCR to diagnose pediatric community-acquired pneumonia. J Infect Chemother 14: 424–432.
43. Cevey-Macherel M, Galetto-Lacour A, Gervaix A, Siegrist C-A, Bille J, et al. (2009) Etiology of community-acquired pneumonia in hospitalized children based on WHO clinical guidelines. Eur J Pediatr 168: 1429–1436.
44. Cilla G, Oñate E, Perez-Yarza EG, Montes M, Vicente D, et al. (2008) Viruses in community-acquired pneumonia in children aged less than 3 years old: High rate of viral coinfection. J Med Virol 80: 1843–1849.
45. Tsolia MN, Psarras S, Bossios A, Audi H, Paldanius M, et al. (2004) Etiology of community-acquired pneumonia in hospitalized school-age children: evidence for high prevalence of viral infections. Clin Infect Dis 39: 681–686.
46. Juvén T, Mertsola J, Waris M, Leinonen M, Meurman O, et al. (2000) Etiology of community-acquired pneumonia in 254 hospitalized children. Pediatr Infect Dis J 19: 293–298.
47. Moore DP, Klugman KP, Madhi SA (2010) Role of *Streptococcus pneumoniae* in

hospitalization for acute community-acquired pneumonia associated with culture-confirmed Mycobacterium tuberculosis in children: a pneumococcal conjugate vaccine probe study. Pediatr Infect Dis J 29: 1099–04.
48. Connell TG, Zar HJ, Nicol MP (2011) Advances in the diagnosis of pulmonary tuberculosis in HIV-infected and HIV-uninfected children. J Infect Dis 204 Suppl : S1151–S1158.
49. Lin PL, Flynn JL (2010) Understanding latent tuberculosis: a moving target. J Immunol 185: 15–22.
50. Oliwa JN, Karumbi JM, Marais BJ, Madhi SA, Graham SM (2015) Tuberculosis as a cause or comorbidity of childhood pneumonia in tuberculosis-endemic areas: a systematic review. Lancet Respir Med 3: 235–243.

CHAPTER 2

Evaluation Of Nasopharyngeal Swab-Types For The Recovery Of *Streptococcus pneumoniae*

2.1 INTRODUCTION

Streptococcus pneumoniae colonizes the nasopharynx of many healthy children and causes pneumonia in a small proportion of those colonized (1-4). Since nasopharyngeal (NP) colonization by *S. pneumoniae* is key to understanding the pathogenesis of pneumococcal disease and is increasingly used as an endpoint for pneumococcal vaccine studies, it is important to establish the optimal strategy for recovery *of S. pneumoniae* from NP specimens. Sampling of the nasopharynx may be achieved using an NP swab or aspirate. NP swabs are preferred as the procedure is simpler, quicker and better tolerated by children. The ideal swab would be highly absorbent, maintain the viability of microorganisms present, release maximum specimen material into culture broth or transport medium and not inhibit downstream applications such as culture and nucleic acid amplification.

The earliest microbiological transport swabs from the 19th century were constructed by winding cotton pledgets around the tip of metal wires or wood, enclosing these in test tubes and sterilizing them in a hot air sterilizer (5). Unfortunately, these swab sticks or wires were often too thick to pass into narrow cavities such as ear canal or urethra and at times too rigid to reliably enter and sample the nasopharynx. It was not until the 20th century that further technological improvements in swab design were made, particularly by the move from wrapped cotton fibre, to the use of calcium alginate swabs in glass or plastic tubes with and without various transport media. The cotton material originally used contained fatty acids which were shown to inhibit bacteria such as Group A Streptococci, *Haemophilus influenzae*, *Corynebacterium diphtheriae*, and *Neisseria gonorrhoeae* (6). Calcium alginate tips on the other hand were a welcomed option for the collection of microbial specimens for analysis. Similar to cotton swabs,

calcium alginate swabs are constructed by wrapping calcium alginate material onto twisted metal or flexible metal or plastic wires. However calcium alginate swabs proved toxic to tissue cultures and inhibited the Polymerase Chain Reaction (PCR) (7-9). Subsequently, non-toxic synthetic fibre-wrapped swabs such as rayon (viscose), manufactured from regenerated cellulose, and polyethylene terephthalate (PET), a thermoplastic polymer resin of the polyester family, manufactured under the brand name Dacron or Terylene, have been developed and has become the preferred swab types utilized in routine microbiological specimen collection products.

Traditional fibre-wrapped swabs, such as PET and rayon swabs are constructed by winding the respective polymer onto the tip of a swab shaft. When stretched from end to end, the length of polyester material extends to over 2 kilometres in length. This creates a deep and hydrophilic central core, which, while absorbing liquid samples very well, does not allow for adequate sample release. Such a swab design may potentially trap a large proportion of clinical material in the fibre matrix, potentially reducing the recovery of microorganisms. Another element for consideration in swab design is to create a swab with a soft and absorbent tip that will cause minimal discomfort or injury to patients during the specimen collection process. Ideally, a swab should allow collection of specimens from various orifices and anatomical sites - a “one size fits all” approach, enabling a single swab type to be used for collection of various samples for multiple downstream applications

To address the limitations of sample release from the swab once absorbed, synthetic thermoplastic polymers, such as nylon, are electrostatically attached to a solid moulded plastic applicator in a perpendicular arrangement by a process called flocking where the fibres are sprayed onto the tip of the shaft, while it is held in an electrostatic field. Nylon

flocked swabs have been shown to improve specimen collection through efficient specimen material release from the swab when compared to spun polymer swabs (10). Flocked swabs absorb liquid by a capillary action that is created by the surface tension between the fibre strands. This brush-like design allows effective dislodging and retrieval of cellular material. Unlike the traditional fibre wrapped swab design, flocked swabs have no inside core to entrap clinical samples or analyte (Figure 1). The short flocked perpendicular fibres create an open structure unlike traditional wrapped swabs and the hydraulic forces between the perpendicular strands are passive ensuring both rapid and spontaneous release of the analyte when placed in contact with a microscope slide, culture plate, preservation medium or assay cartridge.

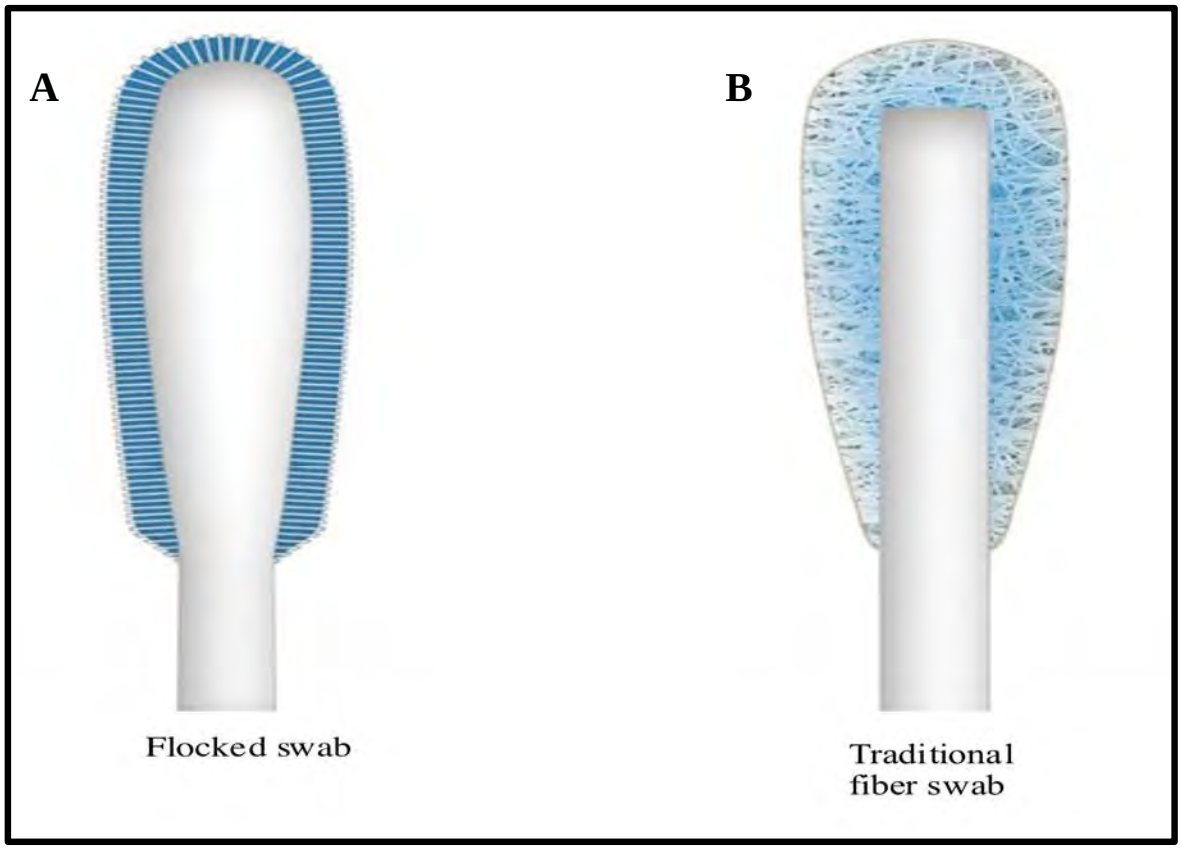

Figure 1: Differences in nasopharyngeal swab design. The brush-like design of nylon flocked (A) swabs allows for more efficient absorption and release of analyte compared to the fibre matrix design of traditional swabs (B) such as PET (Dacron™) and rayon swabs (Adopted from http://www.rapidmicrobiology.com/)

Limited data exist on the recovery of *S. pneumoniae* using different types of swabs (11). Rubin *et al*. showed that rayon swabs were superior to both calcium alginate and PET swabs for the recovery of *S. pneumoniae* (12). To date, there are no published data on comparative recovery of *S. pneumoniae* from flocked swabs.

2.2 Aims and Objectives

The aim of the experiments described in this chapter is to compare pneumococcal recovery from nylon flocked, polyethylene terephthalate (PET) and rayon swabs by

quantitative culture and lytA-targeted real-time polymerase chain reaction (qPCR). The specific objectives are:

1. To compare the liquid volume absorption by nylon flocked, PET and rayon swabs.
2. To compare the quantitative culture recovery of *S. pneumoniae* from nylon flocked, PET and rayon swabs employing calibrated cell suspensions.
3. To compare bacterial NP culture recovery employing nylon flocked and PET swabs from healthy volunteers.
4. To compare quantitative nucleic acid amplification of *S. pneumoniae* obtained from nylon flocked and PET swabs from healthy volunteers.

2.3 METHODS

2.3.1 Swab liquid volume absorption

The immersion swab technique was used to compare the volume of liquid that could be absorbed by each swab type (13). Briefly, a 25 ml graduated cylinder was filled with 20 ml PBS and weighed before and after each swab was submerged for 5 seconds. Each experiment was repeated 10 times over three separate days.

2.3.2 Quantitative culture recovery *S. pneumoniae* using mock specimens

2.3.2.1 Inoculum Preparation

Mid-log phase broth culture of *S. pneumoniae* (ATCC 49619, serotype 19F) was identified by inoculating an overnight culture into 10 ml Todd-Hewitt broth and

incubating at 37°C in 5% CO_2 with continuous gentle shaking at 125 rpm. A 1 ml aliquot of broth culture was aspirated hourly over 6 hours and subjected to spectrophotometry at 600 nm (BioDrop Duo, Cambridge, United Kingdom).

Once mid-log was established for *S. pneumoniae* serotype 19F(Figure S1.1), the process was repeated until 0.5 OD_{600nm} was reached (3 hours), the cell suspension was immediately centrifuged at 10,000 rpm, supernatant discarded and the resulting pellet placed on ice. The cell pellet was washed twice in 1 ml 1 X phosphate-buffered saline, pH 7.4 (PBS; Sigma-Aldrich, St. Louis, United States of America) and finally resuspended in 1 ml PBS.

In order to determine the number of *S. pneumoniae* colony forming units (CFUs) at mid-log phase corresponding to 0.5 OD_{600nm}, 10 fold serial dilutions were made of the primary stock culture (Figure 2) in 1 x PBS solution in 1.5 ml conical bottom microcentrifuge tubes (Corning Inc, New York, United States of America). A 20 µl aliquot of the 10^4 to 10^6 dilution was plated in triplicate onto blood agar (BA) plates containing Columbia blood agar base with 2% agar, 5% horse blood and 4 mg/ml gentamicin (Greenpoint Media Laboratory, National Health Laboratory Service, Cape Town, South Africa), and incubated overnight at 37°C in 5% CO_2. The number of CFUs was determined to be 1.2 x 10^8 CFU/ml (Table S1), which corresponds to 0.5 OD_{600nm} at mid-log phase. Following quantification, the harvested cells were stored at -80°C in 1 ml of skim milk-tryptone-glucose-glycerol (STGG) transport medium, described by O'Brien *et al*., and supplied by Wits Health Consortium, Johannesburg, South Africa (14). To confirm that the viability of the cells was not compromised during storage, quantification of CFUs was repeated after 24 hours of storage. The above experimental procedure was repeated for *S. pneumoniae* serotype 1 and 5 (Table S2.1).

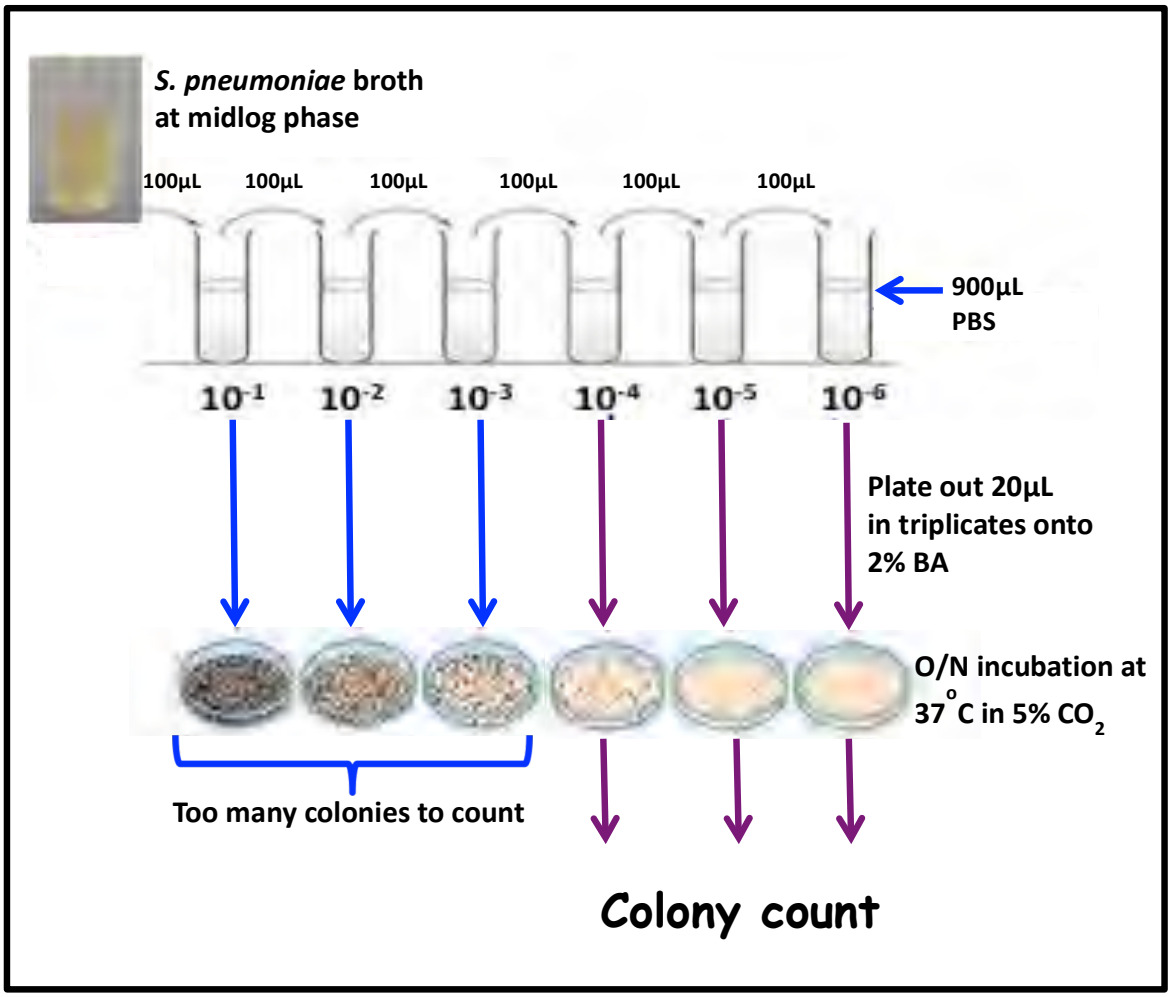

Figure 2: *S. pneumoniae* dilution series for the quantification of the viable cell count. Each dilution was plated out in triplicates as indicated by the purple arrows. The colonies from the first three dilutions (blue arrows) were not included in colony counts mostly due to the dense growth.

2.3.2.2 Quantitative culture recovery of *S. pneumoniae* from nylon flocked, PET and rayon tipped NP swabs using mock specimens

A standardized inoculum of a reference strain of *S. pneumoniae* (ATCC 49619, serotype 19F) and two clinical isolates (serotypes 1 and 5) were cultured for 3 hours to mid-log phase of 0.5 OD_{600nm} in Todd-Hewitt broth from which 10-fold serial dilutions were prepared in PBS (see section 2.3.2.1). Of the dilutions made, 10 aliquots of 20 µl from

the 10^{-2} to 10^{-4} dilutions were transferred to 200 µl reaction tubes in triplicate for each of the three *S. pneumoniae* strains, respectively.

Three different types of NP swabs were compared, the nylon flocked swab (cat. no. 516C; Copan Italia, Brescia, Italy), PET swab (Dacron®) (cat. no. MW151D; Medical Wire & Equipment, Corsham, United Kingdom) and rayon swab (cat. no. 160C; Copan Italia, Brescia, Italy). A single swab of each type was placed into triplicate aliquots of the three *S. pneumoniae* strains at 10^{-2} to 10^{-4} dilutions respectively for 10 minutes at room temperature until the inoculum was absorbed into the swab. The inoculum containing swabs were then transferred into 1 ml STGG. After 30 minutes, the vials containing STGG and swabs were vortexed for 15 seconds, and 20 µl subsequently inoculated onto BA plates and incubated at 37°C in 5% CO_2 overnight. In order to simulate 100% release of *S. pneumoniae* strains (referred to as control) from an inoculated swab, 20 µl aliquots of the 10^{-2} - 10^{-4} dilutions were directly inoculated into 1 ml vials of STGG from which20 µl aliquots were directly inoculated onto BA plates. The plates were incubated at 37°C in 5% CO_2 overnight and colony-forming units (CFU) determined (Figure 3). The percentage of *S. pneumoniae* recovery was calculated as the proportion of the mean CFU recovered from each swab type divided by the mean CFU recovered from the "control" (simulating 100% CFU recovery). The experiment was repeated over three different days. In order to determine the intra- and inter assay variability of these assays, we used the 1/1000 dilution series for each strain. It was mostly this dilution that we consistently recovered CFUs that could either be counted or where there was growth for all replicates considered (Table S1.1 - 1.5).

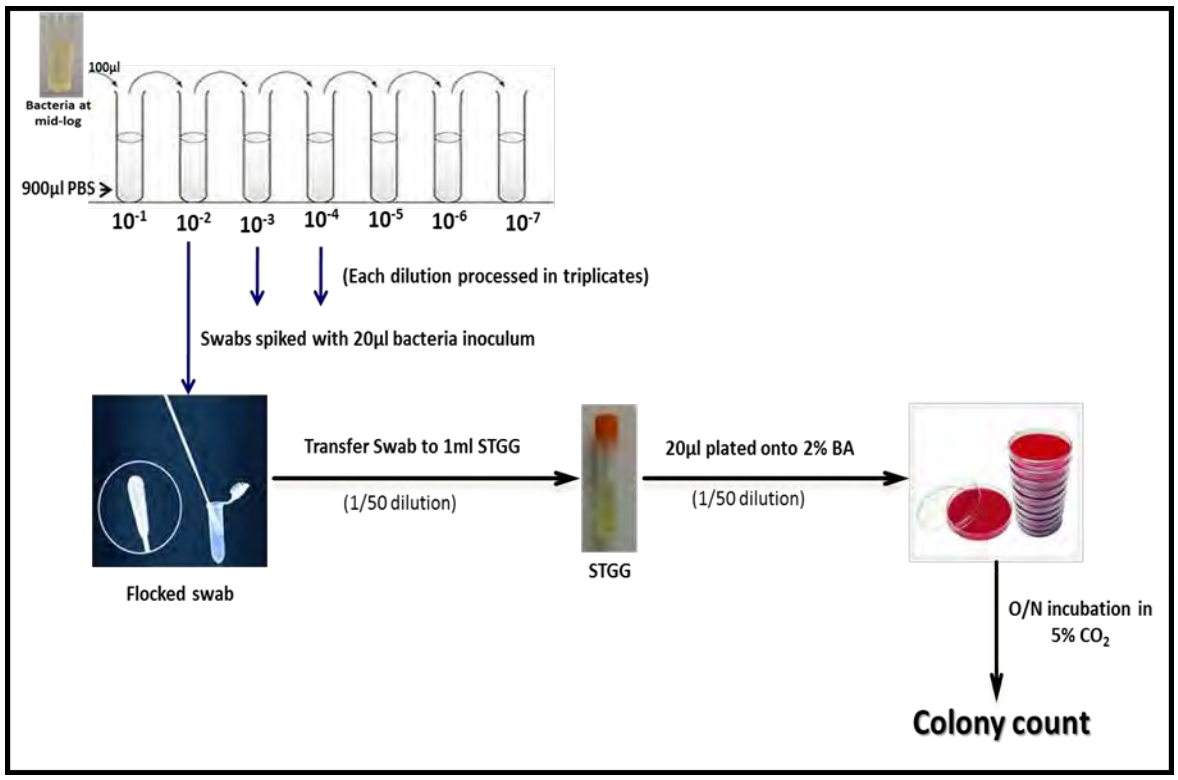

Figure 3: Quantitative culture recovery of *S. pneumoniae* from nylon flocked, PET (Dacron®) and rayon tipped NP swabs. The figure only shows the experimental approach for nylon flocked swabs although repeated similarly for PET and rayon swabs.

2.3.3 Quantitative culture recovery of *S. pneumoniae* from nylon flocked and PET tipped NP swabs from healthy volunteers

Paired nasopharyngeal swabs were obtained from 42 otherwise healthy children undergoing elective surgery at Red Cross War Memorial Children's Hospital. Paired swabs comprised nylon flocked and PET swabs. Rayon swabs were not included as previous findings demonstrated poor recovery of *Staphylococcus aureus* ((15–19). In addition, our own findings (below) had demonstrated that nylon flocked and PET swabs were better than rayon swabs for *S. pneumoniae* release, and there are practical difficulties in performing a 3-way comparison. Paired NP swabs were obtained from

each child from separate nostrils. A nylon-flocked swab was used to obtain a specimen from one nostril and a PET swab from the other nostril. The order of sampling was randomized.

Following sampling, swabs were immediately placed into 1 ml of STGG, transported on ice to the laboratory and frozen at -80^{o}C for later batch processing (Figure 4). After thawing, STGG samples were vortexed for 15 seconds before a 20 µl aliquot was inoculated onto BA plates and incubated at 37^{o}C in 5% CO_2 overnight. Presumptive *S. pneumoniae* colonies were identified by colony morphology, α-hemolysis and ethy-hydrocupreine (optochin) disk susceptibility (Oxoid, Basingstoke, United Kingdom). A quantitative culture was performed by counting the number of colonies to determine the mean CFU of *S. pneumoniae* recovered from each swab type.

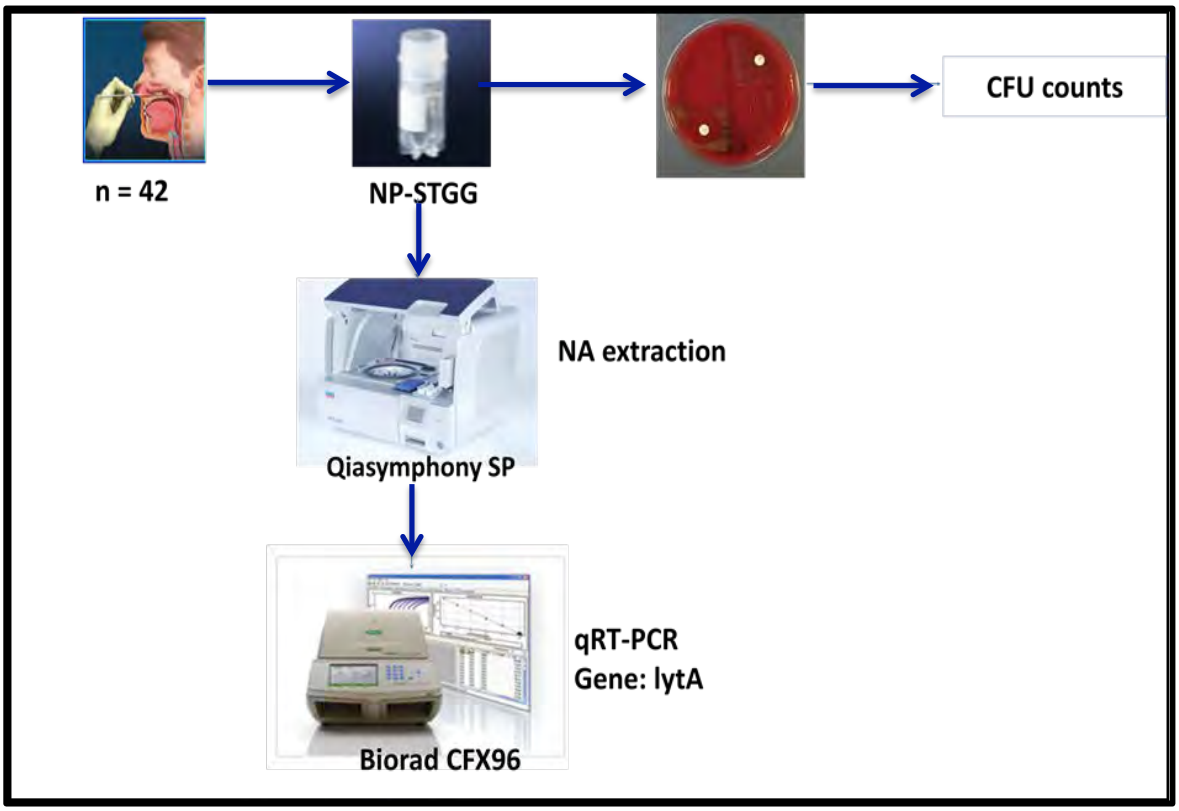

Figure 4: Recovery of *S. pneumoniae* from apparently healthy children. Comparative quantitative cultures and nucleic acid based testing were performed on each paired NP swab.

2.3.4 Comparative nucleic acid amplification of *S. pneumoniae* obtained from nylon flocked and PET tipped NP swabs from healthy volunteers

2.3.4.1 Total nucleic acid extraction

The STGG aliquots obtained from section 2.3.3 were thawed at room temperature and vortexed for 15 seconds. Thereafter, 300 µl of each sample was subjected to automated total nucleic acid extraction employing the QIAsymphony® Virus/Bacteria mini kit on the QIAsymphony SP according to the manufacturer's instructions (Qiagen, Hilden, Germany). Total nucleic acid was extracted from 300 µl STGG aliquots, eluted with 60 µl elution buffer and stored at -20°C prior to nucleic acid amplification.

2.3.4.2 *S. pneumoniae* nucleic acid amplification

Quantitative real-time PCR (qPCR) assay for the detection of the *S. pneumoniae* autolysin-encoding gene (*lytA*) was performed as described by Carvalho *et al* (20). The assay specificity was assessed by screening a pool of heterogeneous bacterial species (Table 2). The reaction mix contained 2.5 µl genomic DNA in a total reaction volume of 12.5 µl containing; 1X TaqMan® gene expression mastermix (Applied Biosystems, California, United States of America) and 200 nM of each primer and probe. PCR amplification was performed on the Bio-Rad CFX96 Touch™ Real-Time PCR amplification system (Bio-Rad Laboratories, Hercules, United States of America). Thermal cycling conditions consisted of an initial hot start of 50°C for 2 minutes, denaturation at 95°C for 10 minutes, followed by 40 amplification cycles of 95°C for 15 seconds and 60°C for 1 minute. The lowest limit of detection of the qPCR assay was 10 copies/ml as determined by inspection of a standard curve (10-fold serial dilution of *S. pneumoniae* ATCC 49619 strain genomic DNA). The quantification cycle (Cq) value at the detection limit point was 36. A positive (*S. pneumoniae* ATCC 49619) and non-template control (sterile water) were included in each qPCR run. Results below the lowest limit of detection were considered negative.

2.3.5 Statistical analysis

Statistical analysis was performed using GraphPad Prism version 6.01 (GraphPad Software Inc, California, United States of America) and STATA software version 11.0 (Stata Corporation, Texas, United States of America). For normally distributed data, unpaired student t-test was used to compare the means of two groups. Wilcoxon rank-

sum test was used to assess the median between two groups when the data was not normally distributed. Analysis of variance was performed to determine whether the mean CFU of *S. pneumoniae* recovered was different within triplicates of each swab type or across *S. pneumoniae* strains when the experiments were repeated on three different days. Analyses exclude the zero values. Including the zero's would indicate a much lower median. The converse is true if the zero values are excluded and only focus on children with a positive swab. To ensure that we captured all of this information we first reported the percentage of children with a positive swab and then calculated the median/mean for those with a positive swab only. A two-tailed p value less than 0.05 was considered as significant.

2.3.6 Ethical considerations

Written informed consent was obtained from a parent or legal guardian. The study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town (HREC ref: 062/2011).

2.4 RESULTS

2.4.1 Swab liquid volume absorption

Nylon flocked swabs absorbed between 70 – 90 µl, PET 20 – 30 µl and rayon 30 - 40 µl PBS. Table 1 shows replicate testing results. There was a significant difference in the amount of liquid absorbed by the nylon flocked, PET and rayon swabs, $p < 0.001$

Table 1: Comparison of liquid volume absorption by nylon flocked, PET and rayon swabs

	Volume absorbed		
Replicate #	flocked	PET	rayon
1	80 µl	30 µl	30 µl
2	80 µl	20 µl	30 µl
3	90 µl	30 µl	40 µl
4	90 µl	30 µl	30 µl
5	90 µl	30 µl	40 µl
6	90 µl	30 µl	30 µl
7	70 µl	30 µl	30 µl
8	80 µl	30 µl	30 µl
9	80 µl	30 µl	40 µl
10	70 µl	30 µl	30 µl
Average (±SD)	82 (±7.89) µl	29 (±3.16) µl	33 (±4.83) µl

2.4.2 Quantitative culture recovery of *S. pneumoniae* from mock specimens

There was no statistical difference in the mean CFU of *S. pneumoniae* recovered within triplicates of each swab type performed on the same day (p = 0.85) or when the experiments were repeated on three different days (p = 0.89), suggesting that the experiments were reproducible. Therefore the *S. pneumoniae* CFU recovered from triplicates of each swab type and on three different days were pooled (Table S2.2).

Overall, there was a significant difference in the mean CFU of *S. pneumoniae* recovered across all the swab types (p < 0.001) and across the three *S. pneumoniae* strains (p = 0.012) tested. The mean number CFU of *S. pneumoniae* recovered from the ATCC 49619 (serotype 19F) strain using nylon flocked swabs was higher when compared with PET (18 x 10^7 CFU/ml vs. 7.3 x 10^7 CFU/ml, p < 0.001) and rayon swabs (18 x 10^7 CFU/ml vs. 1.3 x 10^7 CFU/ml, p < 0.001) (Fig. 5A). Expressed as percentage recovery of the initial inoculum, recovery of *S. pneumoniae* ATCC 49619 (Serotype 19F) strain from flocked swabs was 100%, while it was 41% from PET swabs and 7% from rayon swabs (Table 2

and S1.3-S1.5). Similar results were observed for *S. pneumoniae* serotypes 1 (Fig. 5B) and 5 (Fig. 5C).

Table 2: Recovery of *S. pneumoniae* from mock specimens using flocked, PET (Dacron™) and rayon swabs

S. pneumoniae strain ("Control")[a]	Swab type	n[b]	Mean CFU recovery/ml (SD)[c]	Percentage recovery [d]
Serotype 19F[e] (18×10^7)	Flocked	9	18×10^7 (5.7×10^6)	100 (97.1-103.9)
	PET	9	7.3×10^7 (3.7×10^6)	41 (37.7-45.1)
	rayon	9	1.3×10^7 (2.7×10^6)	7 (4.2-12.5)
Serotype 1 (19×10^7)	Flocked	9	19×10^7 (5.4×10^6)	100 (84.3-105.2)
	PET	9	8.5×10^7 (3.8×10^6)	45 (39.0-47.0)
	Rayon	9	1.0×10^7 (2.1×10^6)	5 (1.2-7.8)
Serotype 5 (5.8×10^7)	Flocked	9	5.8×10^7 (3.2×10^6)	100 (91.3-104.0)
	PET	9	2.6×10^7 (8.2×10^6)	45 (8.6-68.6)
	Rayon	9	0.7×10^7 (1.1×10^6)	12 (9.1-12.0)

SD, Standard deviation

(a) Represents a simulated 100% release of *S. pneumoniae* into skim milk-tryptone-glucose-glycerol (STGG) media and is used as a reference to calculate the percentage of recovery. **(b)** Pooled replicates of each swab type from three independent experiments performed over three different days. **(c)**Mean of colony forming units (CFU) obtained from 9 different replicates of each swab type tested. **(d)**Percentage of recovery was calculated as the proportion of the mean CFU recovered from each swab type divided by the control (100% CFU recovery) for each respective *S. pneumoniae* serotype. **(e)***S. pneumoniae* serotype 19F tested in this study corresponds to the American Type Culture collection (ATCC 49619) strain (21).

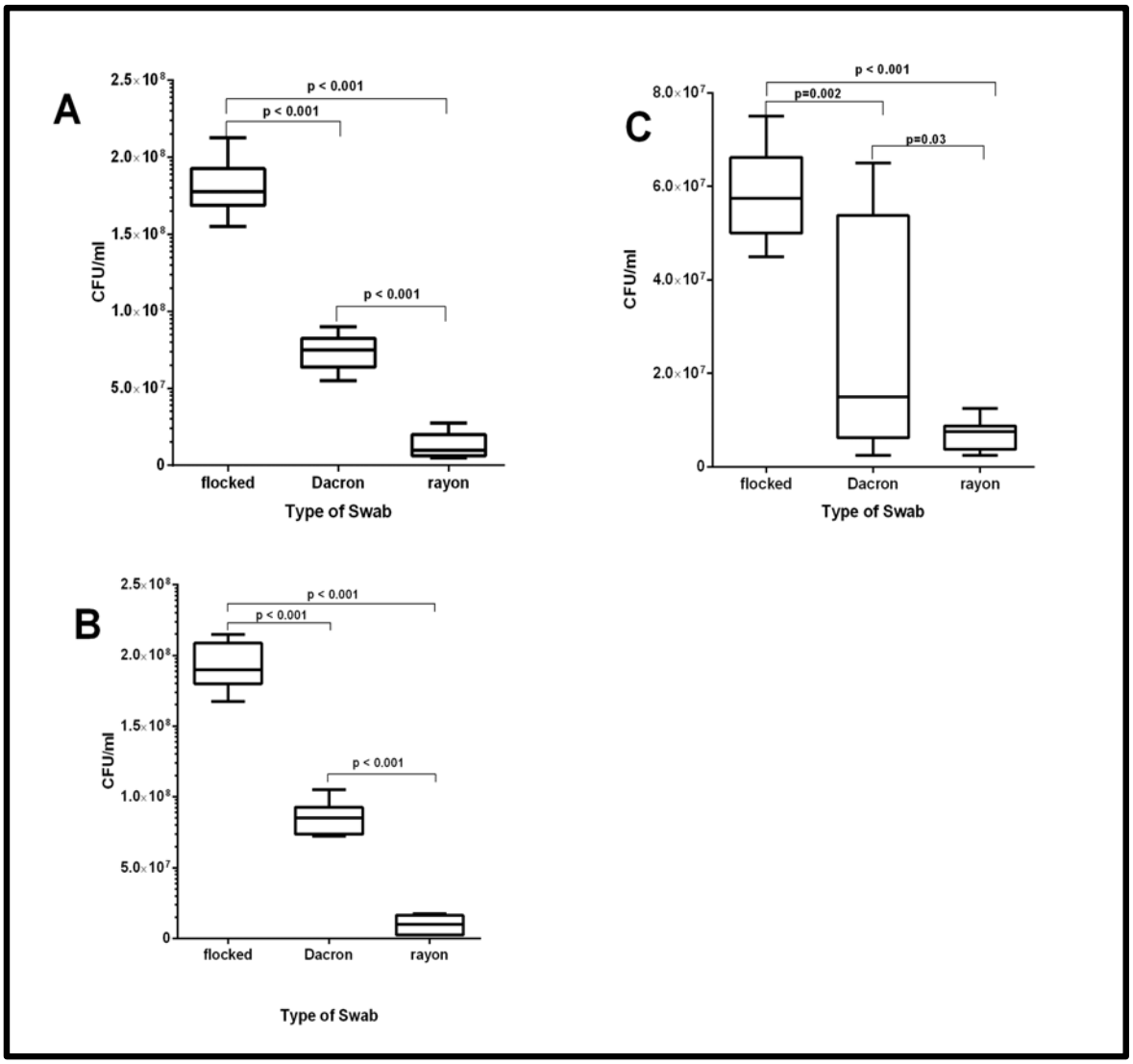

Figure 5: Recovery of *Streptococcus pneumoniae* from mock specimens using nylon flocked, PET (Dacron™) and rayon swabs. The data is presented as the pooled mean CFU recovered from 9 replicates of each swab type using (Tables S1.3-1.5) *S.pneumoniae* serotype 19F (Fig 5A), serotype 1 (Fig 5B) or serotype 5 (Fig 5C). An unpaired student t-test was used for statistical comparisons.

2.4.3 *S. pneumoniae* recovery from NP samples collected from healthy volunteers

2.4.3.1 Quantitative culture recovery of *S. pneumoniae* obtained from nylon flocked and PET tipped NP swabs from healthy volunteers

Paired (flocked and PET) NP samples were collected from 42 healthy children (median age 8 [IQR 5–16] months). *S. pneumoniae* was cultured from 18 of the 42 (43%) NP samples. In 12 of 42 (29%) participants *S. pneumoniae* was recovered from both flocked and PET swabs, whilst in four (10%) and two (5%) participants *S. pneumoniae* was recovered only from flocked and PET swabs, respectively. The children from whom *S. pneumoniae* was recovered were younger than those from whom it was not (8 months vs. 12 months, $p = 0.032$). Although the median number of CFU recovered from flocked swabs was approximately two-fold higher (8.8×10^4 CFU/ml [IQR, 2.0×10^2 – 4.0×10^5 CFU/ml]) than that of PET swabs (3.7×10^4 CFU/ml [IQR, 4.0×10^2 – 3.2×10^5 CFU/ml]), this difference was not significant, $p = 0.17$ (Figure 7B).

2.4.3.2 Comparative nucleic acid amplification of *S. pneumoniae* obtained from nylon flocked and PET tipped NP swabs from healthy volunteers.

2.4.3.2.1 *S. pneumoniae lyt*A qPCR assay specificity

Assay specificity was tested by including other heterogeneous bacterial species (Table 3), some of which include a range of streptococci and common mucosal colonisers. An exponential amplification signal was detected for *S. pneumoniae* including the positive control. None of the other heterogeneous bacterial species were amplified (Table 3 and Figure 6). Amplification of 16S rRNA genes as a positive control in non-pneumococcal strains confirmed the presence of DNA template in these samples (Table 3).

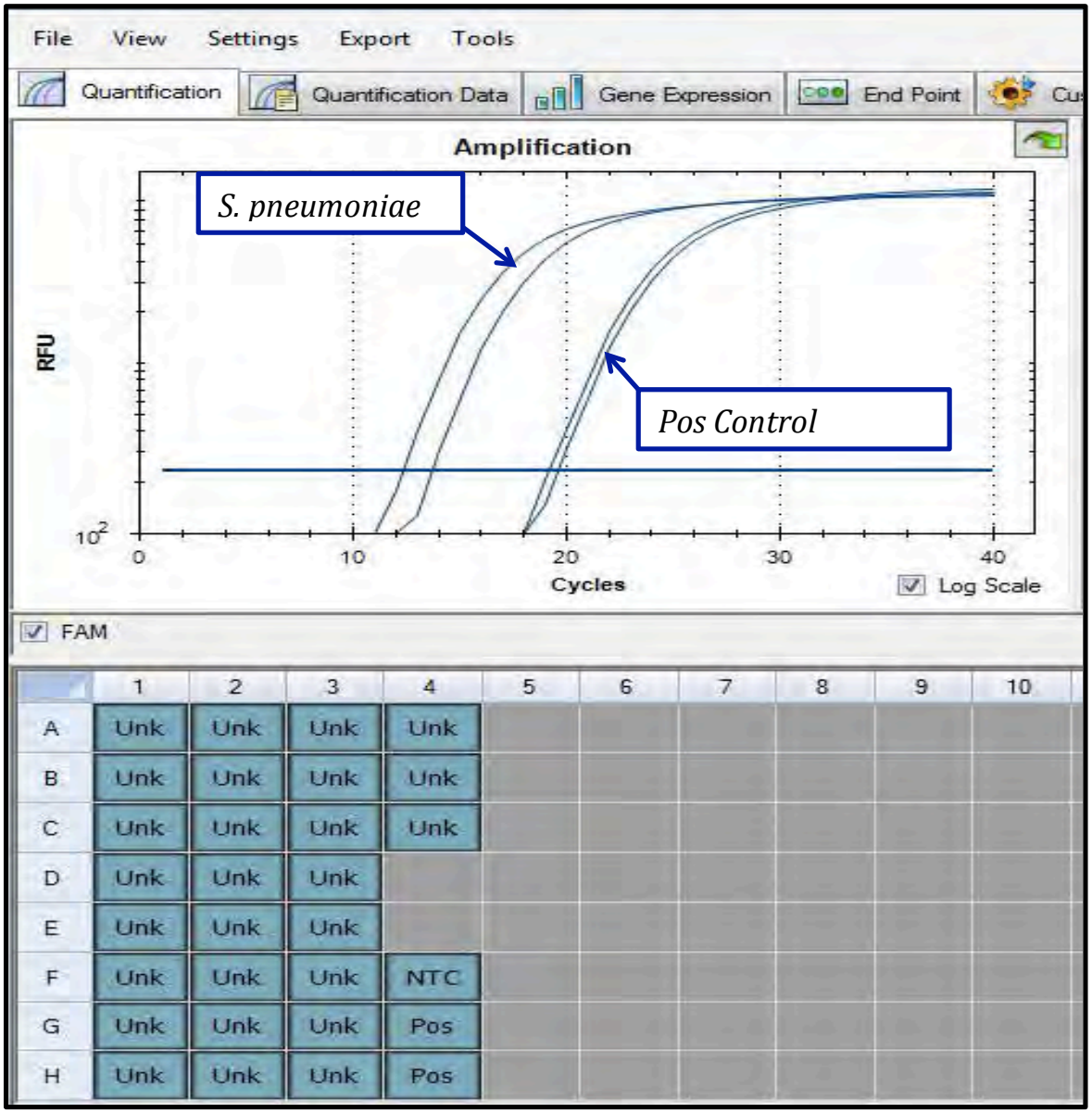

Figure 6: Screenshot of the plate set up and results of *Streptococcus pneumoniae* autolysin (*lyt*A) gene-targeted quantitative real-time PCR. Only *S. pneumoniae* was amplified as seen by the exponential amplification signals.

Table 3: Assessment of *S. pneumoniae* autolysin gene-targeted quantitative real-time PCR assay specificity

Organism	Type strain	*16S* rRNA*	lytA PCR result
Enterococcus faecalis	#*ATCC 29212*	+	-
Enterococcus faecalis	*ATCC 512299*	+	-

Streptococcus pneumoniae	*ATCC 700669*	+	+
Streptococcus pneumoniae	*ATCC 6303*	+	+
Streptococcus pneumoniae	*ATCC 49619*	+	+
Streptococcus anginosus	*CAP QC 2008*	+	-
Streptococcus bovis	*ATCC 9809*	+	-
Streptococcus equi	*ATCC 43079*	+	-
Streptococcus pyogenes	*ATCC 12203*	+	-
Streptococcus pygenes	*ATCC 19615*	+	-
Streptococcus pygenes	*ATCC 8668*	+	-
Streptococcus pygenes	*NSCC stock*	+	-
Streptococcus thermophilis	*ATCC 19258*	+	-
Streptococcus viridans	*NCTC 7868*	+	-
Staphylococcus saprophyticus	*ATCC 15305*	+	-
Staphylococcus aureus	*ATCC 25923*	+	-

#ATCC, American Type Culture Collection; NCTC, National Collection of Type Cultures.
CAP QC, Community acquired pneumonia quality control isolate; NSCC, National stock Culture Collection
+, Detected; **-**, Not detected.
*rRNA PCR was performed as described by Woo *et al* (22).

2.4.3.2.2 Detection of S. *pneumoniae* from paired NP swabs

Using qPCR, *S. pneumoniae* was detected in 27 of the 42 (64%) NP samples, with similar detection rates from nylon flocked and PET swabs (26 out of 42 [62%] vs. 24 out of 42 [57%]; $p = 0.657$). *S. pneumoniae* was detected in both flocked and PET swabs in 55% (23 out of 42) of the participants. In three participants, *S. pneumoniae* was detected from nylon-flocked swabs only and in one participant from PET swab only ($p < 0.001$). Comparing the bacterial loads of *S. pneumoniae* detected by *lyt*A qPCR from paired NP samples, the median copy number of S. *pneumoniae* detected from nylon flocked swabs

(3.0 x 10^5 genome copies/ml [IQR, 1.3 x10^2 - 1.8 x 10^6]) was significantly higher than from PET swabs (9.3 x 10^4 genome copies/ml [IQR, 7.0 x 10^1 - 1.1 x 10^6]; p = 0.005) as illustrated in figure 7A.

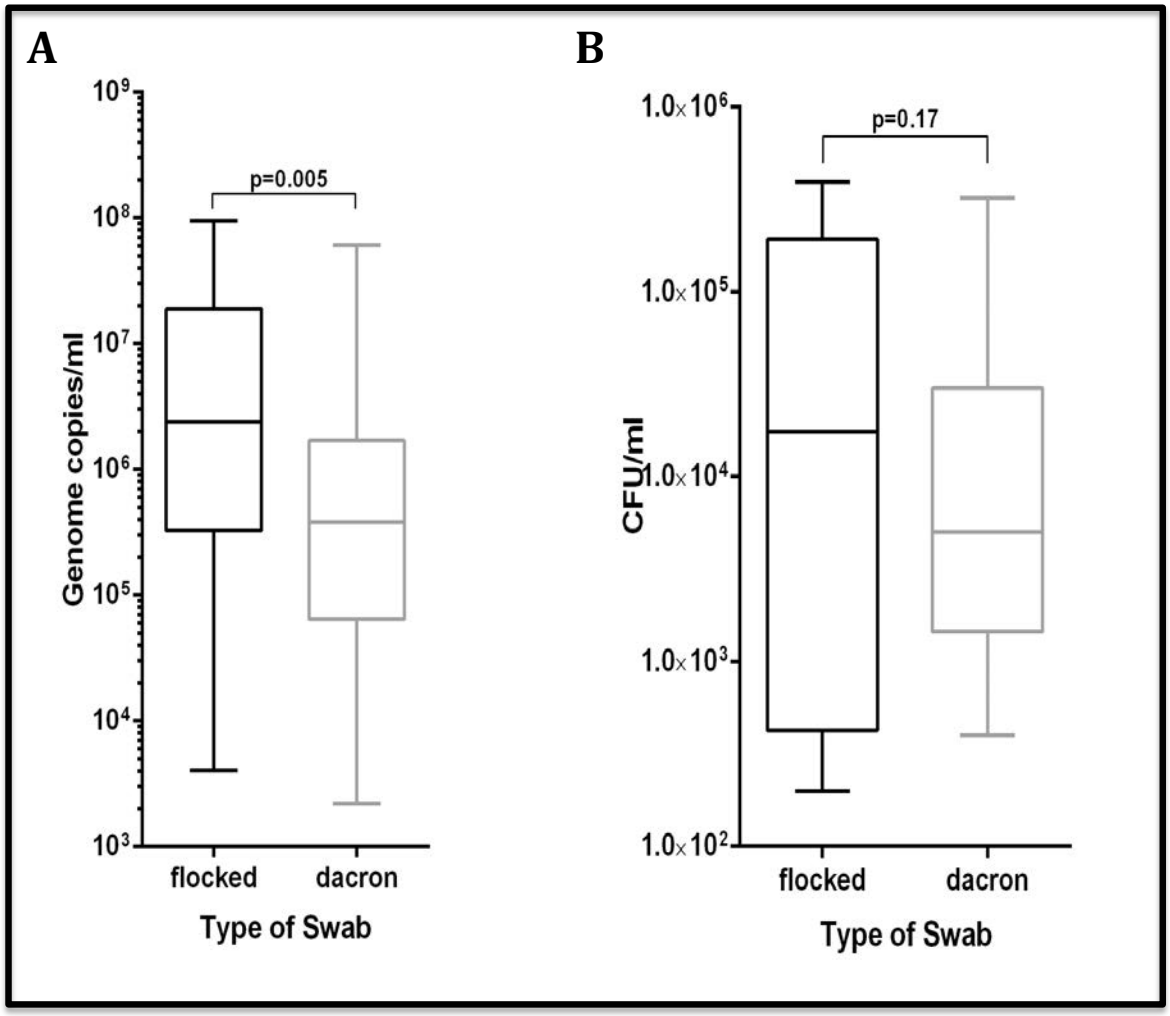

Figure 7: Bacterial load of *Streptococcus pneumoniae* from nasopharyngeal specimens collected using either nylon flocked or PET (Dacron™) swabs from apparently healthy children as determined by both *lyt*A targeted qPCR (A) and quantitative culture (B).

The comparison of *S. pneumoniae* detection by culture and qPCR from NP samples collected from healthy children is summarized in Table 4. The proportion of positive results for *S. pneumoniae* by qPCR was higher when compared to bacterial culture (64% [27/42] vs. 43% [18/42]; p = 0.049).

Table 4: Comparison of *S. pneumoniae* detection by culture and real-time PCR

		Real-time PCR					
		Flocked swab			PET swab		
		Positive	Negative	Total	Positive	Negative	Total
Culture	Pos	16 (38%)	0	16 (38%)	13 (31%)	1 (2%)	14 (33%)
	Neg	10 (24%)	16 (38%)	26 (62%)	11 (26%)	17 (41%)	28 (67%)
Total		26 (62%)	16 (38%)	42 (100%)	24 (57%)	18 (43%)	42 (100%)

PCR, Polymerase chain reaction targeting the autolysin gene (*lytA*) gene of *S. pneumoniae.*

2.5 DISCUSSION

A report from World Health Organisation (WHO) working group (23) and unpublished data from the US Centres for Disease Control and Prevention (CDC) (24) suggests the use of either PET or calcium alginate swabs collected in STGG transport medium for *S. pneumoniae* carriage studies. The present study demonstrates that nylon flocked swabs demonstrate increased absorptive capacity and increased yield of *S. pneumoniae* from mock specimens. In addition, PET swabs recovered significantly more *S. pneumoniae* than rayon swabs. The differences observed in the liquid absorption support the notion that the brush-like design of nylon flocked swabs creates a highly absorbent layer of fibres that absorb liquid by a capillary action (10,13,18). The findings from this study are in line with what has previously been shown for other bacteria (10,25,26). For instance, Verhoeven *et al*. showed that flocked swabs were superior to rayon swabs for the recovery of *Staphylococcus aureus* using culture (19). Similarly, nylon flocked swabs have been shown to improve the recovery of epithelial cells and viruses compared to rayon swabs (18,27,28). Our finding, that PET swabs were better than rayon swabs for the recovery of pneumococcus is in contrast with the finding of Rubin *et al*. (2008), who

followed a similar protocol (12). In our study, we used twice the volume of medium to inoculate swabs (20 µl vs. 10 µl) than that used by Rubin *et al*. in their in-vitro assay.

Whilst we found no significant statistical difference in bacterial detection by culture between nylon flocked and PET swabs from NP samples collected from children (p = 0.17), we demonstrated a trend to increased CFU counts. This was supported by detection of significantly more *S. pneumoniae* DNA by qPCR from nylon-flocked swabs compared to PET swabs (p = 0.005). The importance of bacterial load has been recently demonstrated by Albrich *et al*. showing that the density of NP colonization by *S. pneumoniae* was higher in patients with pneumococcal community-acquired pneumonia compared to control patients using PET swabs for both culture and *lytA*-targeted qPCR (29). Further, using the PCR-based methods, the severity of pneumonia caused by *S. pneumoniae* has been shown to be associated with an increased bacterial load in two independent studies using serum/blood samples (30,31).

A limitation of the present study was the use of a single nucleic acid extraction platform. It has been shown that the DNA extraction methods may influence the DNA recovery from both cotton and flocked swabs used for the collection of DNA from saliva stains (32). Therefore, in order to assess whether the observed high bacterial load in this study from flocked swabs is not inherent to the DNA extraction method, studies using different nucleic acid extraction methods are warranted. Whilst some concerns may arise due to the NP swabs that were collected from either right or left nostril in a random order, Abdullahi *et al.* showed no evidence of systematic of bias to swabbing either the right or left nostril (p = 0.18) (36). The order of swab type was randomised each time that a nostril was sampled to minimize further bias.

In summary, our findings suggest that flocked swabs may offer improved sensitivity for the recovery and/or detection of *S. pneumoniae* from NP swabs. Importantly, flocked swabs are increasingly used for NP sampling for detection of respiratory viruses by nucleic acid amplification (28,33–35). The use of a single swab type for both viral and bacterial studies would simplify specimen collection protocols (17).

2.6 REFERENCES

1. Bogaert D, De Groot R, Hermans PWM (2004) *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis 4: 144–154.
2. Coles CL, Sherchand JB, Khatry SK, Katz J, Leclerq SC, et al. (2009) Nasopharyngeal carriage of S. pneumoniae among young children in rural Nepal. Trop Med Int Health 14: 1025–1033. doi:10.1111/j.1365-3156.2009.02331.x.
3. Antonio M, Dada-Adegbola H, Biney E, Awine T, O'Callaghan J, et al. (2008) Molecular epidemiology of pneumococci obtained from Gambian children aged 2-29 months with invasive pneumococcal disease during a trial of a 9-valent pneumococcal conjugate vaccine. BMC Infect Dis 8: 81.
4. Hill PC, Akisanya A, Sankareh K, Cheung YB, Saaka M, et al. (2006) Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers. Clin Infect Dis 43: 673–679. doi:10.1086/506941.
5. Councilman WT (1893) The Pathology and diagnosis of diphtheria. Am J Med Sci 106: 540–552.
6. Ellner PD, Ellner CJ (1966) Survival of bacteria on swabs. J Bacteriol 91: 905–906.
7. Crane LR, Gutterman PA, Chapel T, Lerner AM (1980) Incubation of swab materials with herpes simplex virus. J Infect Dis 141: 531.
8. Bettoli EJ, Brewer PM, Oxtoby MJ, Zaidi AA, Guinan ME (1982) The role of temperature and swab materials in the recovery of herpes simplex virus from lesions. J Infect Dis 145: 399.
9. Wadowsky RM, Laus S, Libert T, States SJ, Ehrlich GD (1994) Inhibition of PCR-based assay for Bordetella pertussis by using calcium alginate fiber and aluminum shaft components of a nasopharyngeal swab. J Clin Microbiol 32: 1054–1057.
10. Dalmaso G, Bini M, Paroni R FM (2008) Qualification of high-recovery, flocked swabs as compared to traditional rayon swabs for microbiological environmental monitoring of surfaces. PDA J Pharm Sci Technol 62: 191–199. doi:191-9.
11. Van Horn KG, Audette CD, Sebeck D, Tucker K a (2008) Comparison of the Copan ESwab system with two Amies agar swab transport systems for maintenance of microorganism viability. J Clin Microbiol 46: 1655–1658. doi:10.1128/JCM.02047-07.
12. Rubin LG, Rizvi A, Baer A, Al RET, Icrobiol JCLINM (2008) Effect of Swab Composition and Use of Swabs versus Swab-Containing Skim Milk-Tryptone-Glucose-Glycerol (STGG) on Culture- or PCR-Based Detection of *Streptococcus pneumoniae* in Simulated and Clinical Respiratory Specimens in STGG Transport Medium . Society 46: 2635–2640. doi:10.1128/JCM.01903-07.
13. Warnke P, Warning L, Podbielski A (2014) Some are more equal - a comparative study on swab uptake and release of bacterial suspensions. PLoS One 9: e102215.
14. O'Brien KL, Bronsdon MA, Dagan RON, Yagupsky P, Janco J, et al. (2001) Evaluation of a Medium (STGG) for Transport and Optimal Recovery of *Streptococcus pneumoniae* from Nasopharyngeal Secretions Collected during Field Studies. J Clin Microbiol 39: 1021–1024. doi:10.1128/JCM.39.3.1021.

15. De Silva S, Wood G, Quek T, Parrott C, Bennett CM (2010) Comparison of flocked and rayon swabs for detection of nasal carriage of *Staphylococcus aureus* among pathology staff members. J Clin Microbiol 48: 2963–2964. doi:10.1128/JCM.01617-09.

16. Drake C, Barenfanger J, Lawhorn J, Verhulst S (2005) Comparison of Easy-Flow Copan Liquid Stuart ' s and Starplex Swab Transport Systems for Recovery of Fastidious Aerobic Bacteria Comparison of Easy-Flow Copan Liquid Stuart ' s and Starplex Swab Transport Systems for Recovery of Fastidious Aerobic Bacteria: 3–6. doi:10.1128/JCM.43.3.1301.

17. Turner P, Po L, Turner C, Goldblatt D, Nosten F (2011) Detection of Respiratory Viruses by PCR Assay of Nasopharyngeal Swabs Stored in Skim Milk-Tryptone-Glucose-Glycerol Transport Medium †. Society 49: 2311–2313. doi:10.1128/JCM.00224-11.

18. Daley P, Castriciano S, Chernesky M, Smieja M (2006) Comparison of flocked and rayon swabs for collection of respiratory epithelial cells from uninfected volunteers and symptomatic patients. J Clin Microbiol 44: 2265–2267. doi:10.1128/JCM.02055-05.

19. Verhoeven P, Grattard F, Carricajo A, Pozzetto B, Berthelot P (2010) Better detection of *Staphylococcus aureus* nasal carriage by use of nylon flocked swabs. J Clin Microbiol 48: 4242–4244. doi:10.1128/JCM.01425-10.

20. Carvalho M da GS, Tondella ML, McCaustland K, Weidlich L, McGee L, et al. (2007) Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. J Clin Microbiol 45: 2460–2466.

21. Leung MH, Bryson K, Freystatter K, Pichon B, Edwards G, et al. (2012) Sequetyping: Serotyping *Streptococcus pneumoniae* by a Single PCR Sequencing Strategy. J Clin Microbiol 50: 2419–2427. doi:10.1128/JCM.06384-11.

22. Woo PC, Cheung EY, Leung K, Yuen K (2001) Identification by 16S ribosomal RNA gene sequencing of an *Enterobacteriaceae* species with ambiguous biochemical profile from a renal transplant recipient. Diagn Microbiol Infect Dis 39: 85–93.

23. O'Brien KL, Nohynek H (2003) Report from a WHO Working Group: standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*. Pediatr Infect Dis J 22: e1–e11. doi:10.1097/01.inf.0000049347.42983.77.

24. CDC (http://www.cdc.gov/ncidod/biotech/files/pcr-pneumo-carriage-march2010. Accessed 15 February 2013) *Streptococcus pneumoniae* carriage study protocol.

25. Hedin G, Rynbäck J, Loré B (2010) New technique to take samples from environmental surfaces using flocked nylon swabs. J Hosp Infect 75: 314–317. doi:10.1016/j.jhin.2010.02.027.

26. Van Horn KG, Audette CD, Tucker K a, Sebeck D (2008) Comparison of 3 swab transport systems for direct release and recovery of aerobic and anaerobic bacteria. Diagn Microbiol Infect Dis 62: 471–473. doi:10.1016/j.diagmicrobio.2008.08.004.

27. Abu-Diab A, Azzeh M, Ghneim R, Ghneim R, Zoughbi M, et al. (2008) Comparison between pernasal flocked swabs and nasopharyngeal aspirates for detection of common respiratory viruses in samples from children. J Clin Microbiol 46: 2414–

2417. doi:10.1128/JCM.00369-08.

28. Hernes SS, Quarsten H, Hagen E, Lyngroth a L, Pripp a H, et al. (2011) Swabbing for respiratory viral infections in older patients: a comparison of rayon and nylon flocked swabs. Eur J Clin Microbiol Infect Dis 30: 159–165. doi:10.1007/s10096-010-1064-2.

29. Albrich WC, Madhi S a, Adrian P V, van Niekerk N, Mareletsi T, et al. (2012) Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. Clin Infect Dis 54: 601–609. doi:10.1093/cid/cir859.

30. Werno AM, Anderson TP, Murdoch DR (2012) Association between pneumococcal load and disease severity in adults with pneumonia. J Med Microbiol 61: 1129–1135. doi:10.1099/jmm.0.044107-0.

31. Rello J, Lisboa T, Lujan M, Gallego M, Kee C, et al. (2009) Severity of pneumococcal pneumonia associated with genomic bacterial load. Chest 136: 832–840. doi:10.1378/chest.09-0258.

32. Brownlow RJ, Dagnall KE, Ames CE (2012) A comparison of DNA collection and retrieval from two swab types (cotton and nylon flocked swab) when processed using three QIAGEN extraction methods. J Forensic Sci 57: 713–717. doi:10.1111/j.1556-4029.2011.02022.x.

33. Esposito S, Molteni CG, Daleno C, Valzano A, Cesati L, et al. (2010) Comparison of nasopharyngeal nylon flocked swabs with universal transport medium and rayon- bud swabs with a sponge reservoir of viral transport medium in the diagnosis of paediatric influenza. System: 96–99. doi:10.1099/jmm.0.015305-0.

34. Smieja M, Castriciano S, Carruthers S, So G, Chong S, et al. (2010) Development and evaluation of a flocked nasal midturbinate swab for self-collection in respiratory virus infection diagnostic testing. J Clin Microbiol 48: 3340–3342. doi:10.1128/JCM.02235-09.

35. Debyle C, Bulkow L, Miernyk K, Chikoyak L, Hummel KB, et al. (2012) Comparison of nasopharyngeal flocked swabs and nasopharyngeal wash collection methods for respiratory virus detection in hospitalized children using real-time polymerase chain reaction. J Virol Methods 185: 89–93. doi:10.1016/j.jviromet.2012.06.009.

36. Abdullahi O, Wanjiru E, Musyimi R, Glass N, Scott JAG. Validation of nasopharyngeal sampling and culture techniques for detection of Streptococcus pneumoniae in children in Kenya. J Clin Microbiol [Internet]. 2007 Oct [cited 2016 May 13];45(10):3408–10.

CHAPTER 3

Comparison Of A Multiplex Real-Time PCR And Sequetyping Assay For *Streptococcus pneumoniae* Serotyping

3.2 INTRODUCTION

Pneumococcal serotype identification is essential for disease surveillance and monitoring pneumococcal vaccine effectiveness and serotype replacement. Serotyping of the pneumococcal polysaccharide capsule, the immunogenic component of current vaccines, remains the cornerstone of pneumococcal strain characterization. To date, more than 98 capsular serotypes have been described and new ones continue to be described (1,2). Colonisation of the nasopharynx with a homologous strain of pneumococci precedes the development of invasive and respiratory tract disease (3-5). Multiple pneumococcal serotypes can colonize the nasopharynx successively over long period of time, or at any one time (7-8). Invasive disease is commonly regarded as resulting from a single serotype. Public health programs employ serotype prevalence data from invasive disease to assist vaccine selection.

One of the challenges for pneumococcal surveillance programs is the lack of an inexpensive and robust pneumococcal serotyping method. Assessment of serotyping methods needs to take into account the number of pneumococci to be serotyped, serotype coverage, sample type to be used, how long it takes to perform the test and resource availability.

Pneumococcal typing techniques and their limitations

The pneumococcus is known for its genetic diversity within the capsulation loci (1) and ability to switch its capsular PS (9-11). Accurate identification of the pneumococcal capsular PS, the immunogenic component of the current vaccines, remains the cornerstone of strain characterization. Public health programs employ serotype

prevalence data from invasive disease to assist vaccine selection. The following section describes phenotypic and genotypic capsular typing tools, with performance characteristics summarized in table 5.

Table 5: Comparison of selected pneumococcal capsular typing tools

Method	Comparison criteria						Key Disadvantage
	Specifity	Number of Serotypes	Time	Batch precess	Multiple serotypes	Direct Detection	
Phenotypic Methods							
Quellung	n/a	91	1-2 days	No	No	No	Typing sera expensive, time-inefficient, requires experience to interpret and error-prone. Not applicable to culture negative clinical samples.
Latex agglutination & Co-agglutination	n/a	91	1 day	No	No	No	Commercial latex reagents expensive. If inhouse, extensive QC required and prone mis-interpretation. Not applicable to culture negative clinical samples.
Immunoblots (Dot blot & colony blot)	96%	15	1 day	Yes	Yes	Yes	Requires extensive optimisation, also prone to mis-interpretation, serotype cross reactivity remains an issue .
ELISA[1] and EIA[2]	98%	23	1 day	Yes	No	Yes	Expensive equipment set-up costs and maintainance. Extensive optimisation required. Costly polyclonal typing serum and serotype-specific monoclonal antibodies.
Flow Cytometry			1 day	Yes	Potentially Yes	Yes	
Luminex/Bioplex	n/a	23	1 day	Yes	Potentially Yes	Yes	
Genotypic Methods							
Multiplex PCR	100%	46 - 62	2-3 hours	Yes	Potentially Yes	Potentially Yes	Not quantitative, carry-over amplicon contamination. Closely related serotypes not descrimnated.
PCR-RFLP[3]	n/a	46	1 day	Yes	No	No	Results interpration subjective due to image-based data. Specificity concerns due to use of non-serotype specific genes to assign a serotype
FAF-mPCR[4]	99.2%	40	2-3 hours	Yes	Potentially Yes	Potentially Yes	Expensive equipment costs
PCR/ESI-MS[5]	84.7%	45	1 day	Yes	Yes	Potentially Yes	
mPCR/RLB[6]	71%	32	1 day	Yes	Yes	Yes	Closely related serotypes not discriminated.
Real-time PCR	98.6%	35	1 day	Yes	Yes	Yes	Closely related serotypes not discriminated, probes expensive, limited multiplexing options.
Nanofluidic RT-PCR	95.7%	50	1 day	Yes	Yes	Potentially Yes	Expensive capital equipment when compare to PCR equipments. limited multiplexing options.
Microarray	95.2%	43	1 day	Yes	Yes	Yes	Equipment and reagents expensive, technical experties needed,
Sequetyping	86%	48	1 day	Yes	n/a	No	May be difficult to completely discriminate between all serotypes. Can not detect multiple serotypes. Specificity concerns due to use of non-serotype specific genes to assign a serotype. Lack of curated database
Target enrichment-based NGS[7]	97%	38	1 day	Yes	Potentially Yes	Yes	Expensive capital set-up cost, bioiformatics expertise required. Primer specificity a concern
Sequence inference from WGS[8]	96%	94*	1 day	Yes	Potentially Yes	No	Can not detect serotpes from clinical specimen, expensive

KEY for Comparison criteria
Specifity = Specificity (compared to conventional methods); **Number of Serotypes** = Refers to the number of serotypes that can be detected by the assay; **Batch precess** = Does the assay allow for batch processing; **Time** = Does the assay allow for batch processing; **Multiple serotypes** = Can the Assay detect multple pneumococcal

serotypes; **Direct Detection** = Can the assay determine capsular type directly from a clinical sample without the need to culture and isolate pneumococcus. [1]ELISA = Enzyme-linked immunosorbent assay , [2]EIA = Competitive enzyme immunoassays, [3]PCR-RFLP = PCR-Restriction Fragment Length Polymorphism, [4]FAF-mPCR = Fluorescent capillary electrophoresis, [5]PCR/ESI-MS = PCR-Electrospray ionisation mass spectrometry, [6]mPCR/RLB = PCR - Reverse line blot hybridisation (mPCR/RLB), [7]NGS = Next generation sequencing, [8]WGS = Whole genome sequence (WGS)

3.2.1 Phenotypic methods

Phenotypic capsular typing tools utilize antiserum to determine serotype/-group specific pneumococcal capsular PS.

3.2.1.1 Quellung reaction

Pneumococcal serotype identification has mostly relied on phenotypic serotyping methods, most notably the Quellung method developed in 1902 (12). The antiserum utilized in this assay is costly; methods employed are labour intensive, and require significant technical expertise and experience. Quellung, or the capsular reaction test, is still the classic reference standard for pneumococcal serotyping. Although there maybe variations in how the test is performed, the underlying principle involves an antigen-antibody reaction between a specific pneumococcal capsular polysaccharide (CPS) and serotype-specific antisera (13-valent pneumococcal conjugate vaccine [PCV-13] serotypes: 4, 6B, 9V, 14, 18C, 19F, 23F, 1, 3, 5, 6A, 19A and 7F) (12,13).

This antigen-antibody reaction results in a change in the refractive index of pneumococcal capsule such that it appears "swollen". The test is normally performed using sequential pooled antisera until a positive reaction is observed. Subsequently, the individual group or capsular serotype-specific antisera included in the antisera pool that gave a positive reaction is then used to determine the serogroup or serotype respectively. When viewed under a microscope (after adding a counter stain such as methylene blue), pneumococcal cells stain dark blue and appear surrounded by a sharply demarcated halo which represents the outer edge of the capsule (12-14).

The performance characteristics of other capsular PS typing tools are normally compared to those of Quellung reaction due to its unsurpassed sensitivity and specificity, summarized in table 5 above (15). The Quellung reaction is however less practical for high throughput capsular PS typing strategy and can not detect multiple serotypes within the same sample. Since one requires a viable bacterium to perform this test, autolysis of the bacteria in culture and prior antibiotic therapy or a low bacterial load might result in culture negative clinical specimens and subsequently a failed Quellung test. Other drawbacks of Quellung testing include the high cost of antisera, labor intensity; subjectivity in interpretation and technical expertise requirements and maintaining a comprehensive pool of antisera can be costly.

In an attempt to overcome the aforementioned limitations with the Quellung reactions, timeous, cheaper, easier alternative serological-based serotyping methods were developed. These methods are summarized below with performance characteristics of some of these methods summarized in table 5 and some of them discussed in detail below. Although some of these methods initially showed a lot of potential, very few of them were actually adopted into routine use due to some inherent limitation.

3.2.1.2 Co-aggulation and latex agglutination

A positive result is reported when visible swelling (agglutination) occurs between pneumococcal capsular PS and serotype-specific antisera. The use of a microscope is not required. Due to the initial cross-reactivity that was reported, chessboard system was introduced and served as a basis for the latex-coated serotyping kit (Pneumotest-Latex). It identifies all the 23 serotypes included in the PPV-23 and an additional 25 cross-reacting serotypes. Although agglutination tests were shown to be accurate, rapid and

required lesser amounts of antisera compared to the Quellung reaction (16), there has been false-positive reports with viridans streptococci (17,18). In addition, commercially sourced latex reagents are expensive and producing these reagents in-house requires extensive QC protocols to minimize misinterpretation of results.

3.2.1.3 Immunoblot assays

Fenoll *et al* and Bogaert *et al* have suggested alternative pneumococcal capsular typing tools, dot blot (19) and colony blot (20) methods respectively. Both methods require heavy pneumococcal suspension that is spotted (blotted) onto nitrocellulose membranes, incubation with serotype-specific antisera and addition of goat anti-rabbit immunoglobulin G (IgG) that is conjugated to peroxidase enzymes. A positive result is seen when a dark colour that is surrounded by a halo is seen after adding a chemical substrate to the membranes. The colony blot method is able to identify other minor serotypes within a sample while the dot blot method cannot and at times difficult to distinguish serotypes within a serogroup. These assays remain time-consuming for high throughput capsular typing. Bronsdon *et al* attempted to modify the dot blot assay by directly placing the nitrocellulose membrane on top of blood agar plates with a fresh growth of pneumococci (21). However, limitations seen with the original Immunoblot assays still apply.

3.2.1.4 ELISA and EIA

Enzyme-linked immunosorbent assay (ELISA) and competitive enzyme immunoassays (EIA) for pneumococcal serotyping directly from serum, NP swabs and urine have been developed (22-24). When compared to the methods described above, these techniques require high antibody titers to improve avidity and even when these are produced in-

house, they would still require extensive QC protocols. They are also time-consuming, results available after several days.

3.2.1.5 Flow cytometry

Flow cytometry allows the simultaneous detection of up to 23 different serotypes (later expanded to 36 serotypes), and resolution of cross-reactive serotypes (25-27). Separation of cells is based on the differences in size, granularity and red-fluorescent labels of the microsphere beads. The method uses capsular PS-specific latex or microsphere beads that are coated with rabbit antisera which are capable of binding to specific capsular PS. Following an overnight incubation with individual pneumococcal lysates, fluorescein (green-fluorescence)-conjugated anti-rabbit antibody is then added to the mixture. Microsphere bead size, red-fluorescence, and green fluorescence are measured in a single flow cytometer run. A correct serotype match is observed by the inhibition of the green-fluorescence when there is a match between the capsular PS on the beads and the capsular PS in the individual pneumococcal lysates.

With the use of Luminex and microsphere technology, flow-cytometry provides rapid and accurate capsular typing of pneumococcal isolates and direct detection of serotypes from clinical specimens and potential for detecting multiple serotypes (28). The initial instrument capital expenditure and maintenance costs make this method less practical for large-scale pneumococcal sero-epidemiology studies.

Although not discussed, other less common phenotypic typing methods include counterimmunoelectrophoresis (CIE) (29), capillary precipitin typing (CPT) (30) and fluorescent antibody technic (FAT) (31).

3.2.2 Genotypic methods

Genotypic typing methods identify the pneumococcal capsular PS on the basis of the nucleotide sequence of capsular genes within the capsulation locus. The publicly available of the more than 90 pneumococcal capsular nucleotide sequences (1,2) has facilitated the development of many genotypic typing tools (8,32–41).

3.2.2.1 Multiplex PCR

PCR assays targeting serotype-specific genes within the capsulation locus have been developed. These have increased sensitivity and specificity and require sequential PCR primer pools specific for commonly encountered disease-causing pneumococcal serotypes (40,41). Most of these conventional multiplex PCR assays have limited multiplexing options and are prone to carry-over amplicon contamination. The detection format is agarose gel-based, therefore subjective and not quantitative.

3.2.2.2 PCR-RFLP

Batt *et al* has described a PCR-Restriction Fragment Length Polymorphism (PCR-RFLP) that is based on visualization of digested PCR fragments (42). Serotype determination is based on the amplicon band patterns. This method is error prone due because it does not utilize serotype specific gene for serotype determination, the similarity of banding patterns is subjective.

3.2.2.3 FAF-mPCR

This is a high-throughput techniques, Fluorescent capillary electrophoresis (FAF-mPCR) (43), that couples fluorescent labeled primers overcomes the limitation of end-product detection of conventional PCR by using fluorescent labeled primers which are separated

by capillary electrophoresis. The unique fluorescent data is then analyzed on and automated fluorescence fragment size analyzer. Initial equipment costs are expensive.

3.2.2.4 PCR/ESI-MS

PCR amplicon generated using serotype-specific primers are subjected to electrospray ionisation mass spectrometry (PCR/ESI-MS) (44). The unique molecular patterns determined from differences in the PCR amplicons base composition are compared to a database of known signatures obtained from known capsulation loci. Performance parameters are shown in table 5. This tool can potentially be used to detect the pneumococcal population structure from the MLST data that can be simultaneously be detected on this platform. As with many derivatives of PCR described above, the initial set up cost for the electrospray mass spectrometer is expensive.

3.2.2.5 mPCR/RLB

An alternative capsular typing method that combines mPCR with reverse line blot hybridisation (mPCR/RLB) has been described (37,45,46). Originally designed to only detect serotypes included in PPV-23, this assay has now been expanded to potentially include all the 90 pneumococcal serotypes (45). The assay the can not differentiate serotypes with shared identical or very similar *wzy* sequences with one or two others in the same or closely related serogroups. These include: 7B/7C/40, 10F/10C, 11B/11C, 15F/15A, 19B/19C, 24F/24A/24B, 25F/25A/38, 28F/28A, 32F/32A, 33B/33D, 35F/47F, 35A/35C/42, and 41F/41A. As a result, further confirmatory testing using serotype-specific antisera may be required.

3.2.2.6 Real-time PCR

A number of rtPCR assays have been developed to provide rapid, more sensitive and potentially quantitative pneumococcal capsular typing tool (39,47,48). These assays can be applied to pneumococcal isolates or culture negative clinical specimens. As with other PCR based assays, closely related serotypes cannot be differentiated. These assays require multiple serotype-specific primer-probe pairs to provide broader serotype coverage and multiplexing options may be limited. PCR with subsequent target detection is prone to amplicon contamination and is more labour intensive than rmPCR. A sequential rmPCR described by Pimenta *et al* (39) allowed the identification of 21 serotypes/-groups including the thirteen serotypes covered by the PCV-13 vaccine and 8 additional key serotypes/-groups likely to be encountered in the Africa. When compared to other rmPCR formats (49), this sequential triplex assay is able to resolve serogroup 6 by differentiating 6A/6B from the emergent 6C serotype that is not included in the current PCV-13 formulation (50). rmPCR obviates the need for amplicon manipulation, is highly sensitive, fast and less labour intensive. PCR assays do not require viable isolates and have the potential to detect multiple serotypes simultaneously (39,48,49,51,52).

3.2.2.7 Nanofluidic RT-PCR

This is a high throughput rtPCR that utilizes nanofluidics (Fluidigm Biomark HD system) (53). This technique utilizes previously described primers and newly designed primers to identify up to 50 different pneumococcal serotypes. The assay uses minute reagent volumes and provides both qualitative and quantitative data for both dominant and low

abundance co-colonization serotypes clinical samples. Nonetheless, primer specificity remains a concern, as closely related serotypes cannot be distinguished.

3.2.2.8 Microarray

Microarray assays utilize serotype-specific probes that target the highly variable glycosyl-transferase genes (*wzx* and *wzy* for most serotypes and *wze* in the case of serotype 3) to identify 43 serotypes including all PPV-23 serotypes (8,54,55). Labeled pneumococcal DNA is added to a microarray chip and binds to the corresponding nucleotide sequence on the array chip. After a run, signal and background intensities of each spot are analyzed using a microarray scanner. The microarray capsular typing approach is rapid and has the capability to detect all pneumococcal serotypes and simultaneously detect other virulence factors and antimicrobial resistance markers from pneumococcal isolates or culture negative clinical samples. Microarrays can also detect serotypes with low abundance in co-colonization clinical samples (8). However, the cost of setting up a microarray and the required high level of technical expertise limits its widespread use, especially in developing countries where pneumococcal disease burdens are high.

3.2.2.9 Capsular Sequence Typing

Elberse *et al.* has described a singleplex PCR reaction using multiple primer sets combined with sequencing of resulting amplicons that is similar to the sequetyping assay (34). This method, capsular sequence typing (CST), utilizes 3-forward and 4-reverse PCR primers in a single PCR reaction, which targets the partial sequence of the *wzh* gene. The resulting amplicon is smaller, 506 bp, compared with the *wzh* amplicon size generated using the sequetyping assay (38). A capsular type (CT) is then assigned

by interrogating a public database (http://www.rivm.nl/mpf/spn/cst/). As with the sequetyping assay, the assay utilizes non serotype-specific genes. Changes in the nucleotide sequence within this regulatory gene may affect the regulation of capsule production but this change may not have any biological relevance in the composition of the pneumococcal capsular PS.

3.2.3.0 Sequetyping

More recently sequetyping, a sequence-based typing method, has been described (38). This method employs partial amplification and sequencing of the capsule biosynthesis protein-tyrosine phosphatase gene, wzh, generating a 1,061 bp amplicon which is subjected to sequencing and comparison to published wzh sequences available via GenBank (http://www.ncbi.nlm.nih.gov/blast/) (38). The sequetyping assay was shown to correctly identify all but two of the PCV-13 vaccine serotypes to serotype level whilst serotypes 7F and 18C were "serogroup-specific". In addition, the sequetyping assay could identify up-to 88% (122/138) of a pool of tested pneumococcal strains representing 46 different serotypes to at-least the correct serotype/-group.

Prior to the work described in this chapter, there were no published head-to-head comparisons of the accuracy of the sequetyping vs. multiplex PCR approaches.

3.2.3.1 Sequence inference from WGS

The high sequence homology in the capsulation locus of pneumococcal serotypes within a serogroups and cross-reactivity with closely related non-pneumococcal polysaccharide genes that are observed especially in respiratory samples posses major limitations to target-based genotypic typing approaches. The increasing availability of

new sequence data and reduced cost of next generation sequencing technologies has taken the quest high throughput capsular typing tools to are era of possibilities (56,57). At present, a number of bioinformatics tools can be applied to whole genome sequence (WGS) data to deduce the pneumococcal capsule type. In addition to making serotype inference, WGS data can be used to analyze the sequence diversity in the entire capsulation loci in a greater detail and the associated evolutionary changes in the pneumococcal population structure (9–11,57,58). Although WGS is becoming affordable, the set up cost in developing countries is still high and high technical expertise in sample preparation together with the bioinformatics capability to extract and analyze high throughput data from this platform makes it less attractive especially in developing countries.

3.2.3.2 Target enrichment based NGS

Liyanapathirana *et al* has describe an alternative method to WGS that employs a target enrichment based next generation sequencing approach to identify 33 pneumococcal serotypes (59). Two primer pairs are used; the first pair distinguishes pneumococcus from other streptococcal species using primer that target pneumococcal specific autolysin (*lyt*A)-encoding gene while the second pair targets a 500 bp serotype specific gene within the capsulation loci. The PCR amplicons are enriched prior to sequencing on a bench-top sequencer. Performance characteristics are shown in table 5 above. As with WGS, this method also requires huge set up cost in highly trained expertise and bioinformatics capability to extract and analyze high throughput data. In addition, the primers used cannot discriminate closely related serotypes.

More practical, higher throughput typing techniques are required for expanding public health laboratory services in many areas of the world to support growing disease control programs and epidemiological surveillance. While the newer phenotypic methods all have their distinct benefits and often surpass Quellung in terms of rapidity and cost, some of the methods require sophisticated and expensive instruments. Given the heterogeneity and recombinogenic nature of pneumococci, capsular typing tools which infer type from DNA sequence, including target enrichment-based next generation sequencing (NGS) and whole genome sequencing (WGS) (60) are attractive newer methods to complement the molecular typing methods discussed above and may also aid in resolving discrepant phenotypic and genotypic findings.

3.3 Aims and Objectives

The overall aim of the experiments described in this chapter is to compare rmPCR and sequetyping assays for suitability for high throughput pneumococcal capsular typing. Specific objectives are:

1. To compare rmPCR and sequetyping methods employing 40 pneumococcal strains Quellung serotyped control strains.
2. To examine the ability of rmPCR and sequetyping assays to detect mixed serotypes.
3. Applying rmPCR and sequetyping to pneumococcal strains isolated from the nasopharynx of 83 healthy infants from the Drakenstein Community, Western Cape.

3.4 MATERIALS AND METHODS

3.4.1 Assay validation

3.4.1.1 Control strain preparation

Isolates comprised 40 Quellung-typed control strains, figure 8, (kindly donated by Dr. Anne von Gottberg, Centre for Respiratory Diseases and Meningitis (CRDM), National Institute for Communicable Diseases (NICD), South Africa [61]). These were randomly selected isolates obtained from our collaborators (NICD) that included vaccine serotypes and other serotypes that are commonly circulating in our setting. These isolates were transported on Dorset egg medium (62), subcultured onto Columbia blood agar base with 2% agar, 5% horse blood and 4 µg/mL gentamicin media (CAG) upon receipt (Green point Media Laboratory of the National Health Laboratory Service, Cape Town, South Africa) and incubated at 37^{o}C in 5% CO_2 overnight. The resulting colonies were inoculated into in 1 ml skim milk-tryptone-glucose-glycerol (STGG) transport medium frozen at -80^{o}C for batch processing.

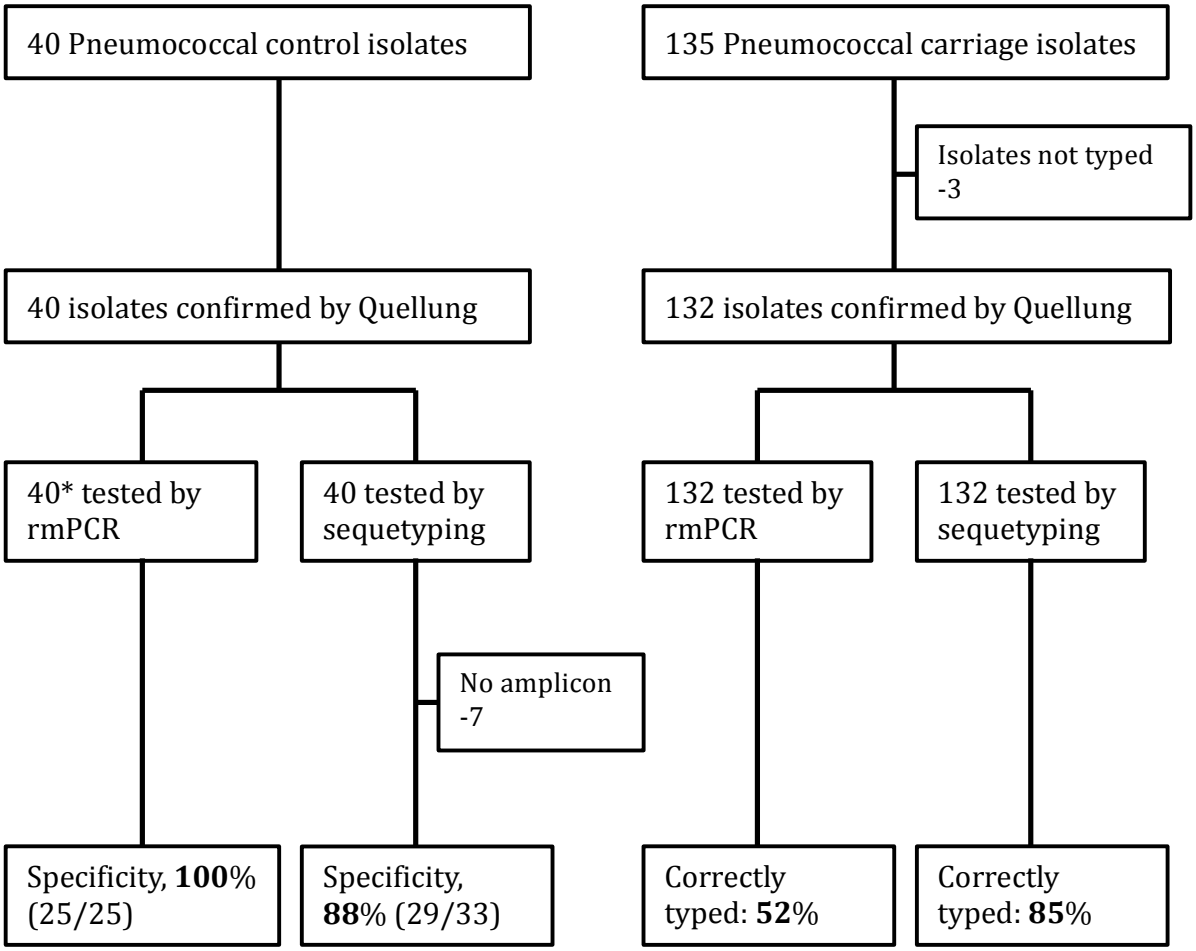

Figure 8: Flow chart showing the pneumococcal isolates that were included in the study. *Of the 40 isolates that were tested by rmPCR, only 25 were included as part of the rmPCR targets.

3.4.1.2 Carriage isolates

Subsequently, 135 pneumococcal isolates (Figure 8) were cultured from nasopharyngeal (NP) swabs that were collected from 83 healthy infants by employing nylon-flocked swabs (Copan Italia, Brescia, Italy). Infants were recruited between May 2012 and September 2013 as part the Drakenstein Child Health Study (DCHS), a South African birth cohort study (63). NP swabs were collected employing the World Health Organization protocol for pneumococcal carriage studies (15). Briefly, the collected NP swabs were immediately placed into 1ml STGG, transported on ice to the laboratory and

frozen at -80°C for batch processing. After thawing, STGG samples were vortexed for 15 s before a 10 µl aliquot was inoculated onto Columbia blood agar base with 2% agar, 5% horse blood (BA) plates and incubated at 37°C in 5% CO2 overnight. Presumptive pneumococcal isolates were identified by colony morphology, α-hemolysis and ethylhydrocupreine (optochin) disk susceptibility (Oxoid, Basingstoke, UK) as previously described (64,65).

3.4.1.3 Nucleic acid extraction

Prior to rmPCR and sequetyping, all isolates were subjected to nucleic acid extraction employing a heat lysis method as previously described (66). Briefly, a sweep of pneumococcal colonies was obtained from primary BA plates that were inoculated with thawed STGG aliquots containing either pneumococcal control strains or carriage isolates. The colony sweeps were resuspended in 100 µl of phosphate-buffered saline, pH 7.4 (PBS; Sigma-Aldrich, St. Louis, MI) thereafter heated at 95°C for 5 min. The supernatant containing genomic DNA (gDNA) was ten-fold serially diluted in PBS before nucleic acid amplification.

3.4.1.4 Real-time multiplex PCR assay

The rmPCR, designed as a 7x3-plex that targets 21 serotypes. We used the multiplex scheme for an African region, based on a relatively limited sample observed, table S3.1 (39). This assay targets all the pneumococcal serotypes included in PCV-13 and 8 other additional serotypes/-groups (PCV-13 serotypes: 1, 3, 4, 5, 6A/6B, 7F/7A, 9V/9A, 14, 18C/18A/18B/18C, 19A, 19F, 23F, 23A; The 8 additional serotypes/-groups include: serotypes 2, 6C/6D, 11A/11D, 12B/2F/46, 15A/15F, 16F, 22F, 33A/33F

Briefly, the PCR reaction comprised 12.5 µl of 2X SensiFAST™ Probe No-ROX One-Step master mix (Bioline, Taunton, MA), primers and probes for serotype as described by Pimenta et al, 5 µl gDNA (diluted 1:1000) and nuclease/RNase-free water (Applied Biosystems, Irving, CA) for a final reaction volume of 25 µl (Table S3.2 – S3.8). The thermal cycling conditions consisted of initial denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and annealing/extension at 60°C for 1 min employing a CFX96 Touch™ Real-Time PCR amplification system (Bio-Rad Laboratories, Hercules, CA)

3.4.1.5 Sequetyping assay

The assay was performed as previously described (38) with minor modification: the PCR reaction comprised 12.5 µl of 2X KAPA T*aq* Ready Mix® (KAPA Biosystems, Boston, MA), 1 µl of primer mix, 2 µl gDNA (diluted 1:10), 8.5 µl nuclease/RNase-free H_2O (Applied Biosystems) in a final volume of 25 µl (Table S3.9). Thermal cycling consisted of an initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 65°C for 30 s, and extension at 72°C for 90 s employing an Applied Biosystems® 2720 Thermal Cycler (Applied Biosystems). The PCR products were separated by electrophoresis in 1.5% agarose gel (SeaKem LE Agarose; Lonza, Rockland, ME) for 45 min at 80 V in a 1X Tris-acetate EDTA buffer. Ethidium bromide-stained DNA products were visualized under UV illumination and sized by using a 1–kb DNA molecular size marker (HyperLadderv1-kb®; Bioline).

PCR products were prepared for sequencing employing Exo-SAP IT® (Affymetrix, Maumee, OH) according to the manufacturer's instructions. Prepared amplicons were submitted for cycle sequencing employing the BigDye® Sequence Terminator kit V3.1 (Applied Biosystems) and analysed on an ABI 3500 XL Genetic Analyzer (Appplied

Biosystems) by Inqaba Biotech (Inqaba Biotechnical Industries [Pty] Ltd, Pretoria, South Africa). Sequencing was performed in both directions using forward (*cps*1), 5'-GCA ATG CCA GAC AGT AAC CTC TAT-3', and reverse (*cps*2), 5'-CCT GCC TGC AAG TCT TGA TT-3' primers.

DNA sequences obtained were assembled and edited using DNA Baser Sequence Assembler v4 (www.DnaBaser.com). The consensus sequences were used to interrogate the GenBank database (http://www.ncbi.nlm.nih.gov/blast/) and assign a serotype using the criteria as per protocol (38). Briefly, the serotype of the *wzh* nucleotide sequence from GenBank with the highest BLAST bit score was assigned, provided that sequence identity was >98% with the query amplicon nucleotide sequence.

3.4.1.6 Sequetyper program

To automate the above process, a Java-based program, Sequetyper (available at http://www.gematics.com/sequetyper.html) was developed and validated to automatically analyse and determine the pneumococcal serotype based on interrogation of GenBank with the input forward and reverse sequences of the generated *wzh* amplicon. This application is suitable for high-throughput analysis of sequetyping data. Briefly, the Sequetyper program will search within a given input directory for the forward and reverse compliment AB1 sequencing trace files using a file-naming convention (detailed in the help menu of the program), perform base calling using a Phred-scaled quality cut-off value of 0.05. Phred-scaled quality scores are assigned as a logarithmic measure of the base-calling error probabilities (Table 6). Phred-scaled quality scores for most sequencing platforms can range from 2 to 63. For many purposes, a Phred-quality score of 20 is acceptable. An example is given below in

figure 9 where a thymine nucleotide base has been assigned a Phred-scaled quality score of 50.

Table 6: Phred quality scores

Phred Quality Score	Error (Probability of incorrect base call)	Base call accuracy (1-Error)
10	1 in 10 = 10%	90%
20	1 in 100 = 1%	99%
30	1 in 1000 = 0.1%	99.9%
40	1 in 10000 = 0.01%	99.99%
50	1 in 100000 = 0.001%	99.999%
60	1 in 100000 = 0.0001%	99.9999%

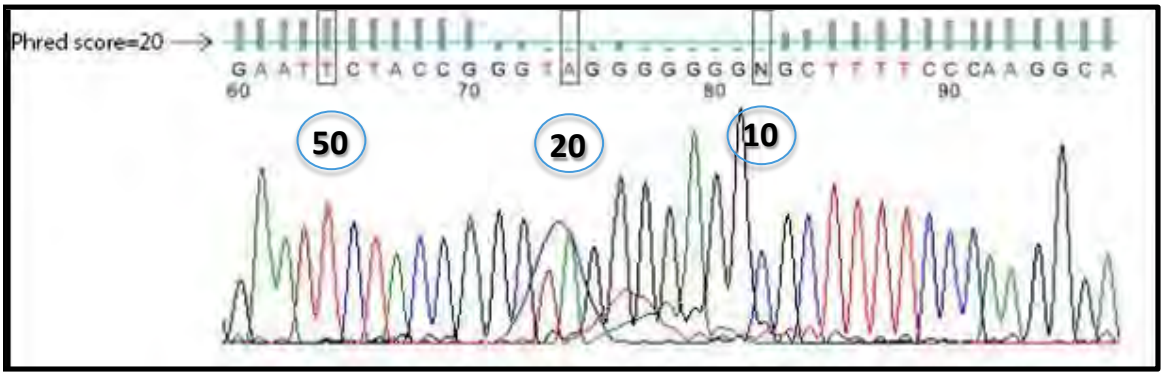

Figure 9: Phred quality scores. The quality of the electropherograph are assesed based on the logarithmic probabilities related to the base-calling error. A high Phred of 50 as in thesse present figure suggests there is one in a 100,000 chances that a Adenine base was called wrong by the sequencing platform.

This implies that there is a 1 in 100,000 chance (99.999% base call accuracy) of incorrectly calling the thymine nucleotide base. The sequence and quality files are read in and the consensus sequences are used to perform a blastn alignment using four stringent criteria. These output files are also generated with the results for review. A typical Phred-quality output for serotype 19F (ATTCC 49619) is shown in figure S3.1. The sequetyper has a local database of nucleotide sequences that are derived from four stringent criteria. The first one is local search within the 92 *cps* sequences made available by the Wellcome Trust (1); the second criterion returns any pneumococcal *cps*

nucleotide sequences; while the third criterion return any pneumococcal nucleotide sequence (This reduces the risk of not picking up any nucleotide sequences that do not indicate the serotype name in its filename) excluding shortgun sequences; and the fourth criterion involves a search for pneumococcal nucleotide sequences from the Capsular Sequence typing database (34). Sequetyper then performs a comprehensive search across all the four criteria, returns a pool of all nucleotide sequences with an identity of >95%. The algorithm then searches within this pool only and best alignment with a nucleotide sequence identity of >98% from the four criteria to assign a serotype. The output files are generated in either ".html", ".xls" or ".txt" format. Because the algorithm uses a local BLAST search, the program can run offline and does not require Internet connectivity. This application is suitable for high-throughput analysis of sequetyping data. Sequetyper was validated employing 175-paired forward- and reverse sequences generated during the study. Briefly, manually identified sequence identities were re-evaluated blindly through sequetyper and results compared. Two manual vs. sequetyper discordant identities were obtained, and subsequent to manual re-evaluation assigned according to original sequetyper results. The discordant results were due to as a result of mislabeling of AB1 files such that the algorithm returned with an error while the second was due to a DNABaser miscalling a nucleotide base when generating a contig.

3.4.2 Discordant Analysis

3.4.2.1 Quellung testing

Pneumococcal control and nasopharyngeal isolates were submitted to CRDM for quellung testing using specific anti-sera (Statens Serum Institute, Copenhagen,

Denmark). Serotype 6C was distinguished from serotype 6A by PCR (67), while serotype 25A and 38 were undistinguishable and hence reported as 25A/38 (68).

3.4.1.2 Next-generation sequencing

Three pneumococcal carriage isolates serotyped as 16F by Quellung and rmPCR but identified as 9V by sequetyping were subjected to WGS. The 3 discordant isolates as well as a control strain identified as 9V by Quellung, sequetyping and rmPCR were also included. Briefly, gDNA was isolated with a Wizard Genomic DNA Purification Kit (Promega Corporation, Fitchburg, WI) according to the manufacturer's instructions. The gDNA quality was assessed using the Qubit® Fluorometer (Life Technologies, Carlsbad, CA), the NanoDrop ND1000 (Life Technologies) and agarose gel electrophoresis used to determine absolute concentration, polyphenolic/polysaccharide/chaotropic salt contamination and gDNA integrity respectively. Quantified gDNA was submitted to the Centre for Proteomics and Genomic Research (CPGR) for WGS. Briefly, sequencing libraries were generated using the Nextera XT DNA Sample Prep Kit (Illumina, San Diego, CA) and the libraries were indexed according to the dual-bar cording protocol (with i7 and i5 primers) using the Nextera XT Index Kit (Illumina). Libraries were then normalized, pooled, and a 5% PhiX control added before sequencing with the Illumina MiSeq Reagent Kit v2 (500 cycle) on the Illumina MiSeq system.

3.4.1.3 *De novo* sequence assembly

The quality of the output sequence data was assessed using FastQC (69) and sequencing adapters were trimmed using Trimmomatic (70). The 3'-end nucleotides with PHRED scores below 20 were trimmed using the fastx_trimmer tool of FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit) (71). The sequence data was then assembled

de novo using SPAdes v3.0.0 assembler (72). Draft genome assemblies were annotated individually using RAST (Rapid Annotation using Subsystem Technology) (73). The contigs containing putative *cps* regions were identified through the standalone *blastall* homology searches against the 16F (Accession: CR931668) and 9V (Accession: CR931648) annotated reference genomes and then extracted to a separate file using a shell command based on SAM tools (74). These contigs were then aligned and visual representation of the alignments was performed using the Artemis Comparison Tool (ACT) v6 and WebACT (75).

3.4.3 Data analysis

Results of the two molecular serotyping assays up to the serogroup level were compared with serotyping results obtained by Quellung testing. In cases of discordance between the two molecular serotyping assays, the results were confirmed by Quellung testing. Serotype distribution was determined based on Quellung results. Where more than one isolate was tested from the same child, isolates of the same serotype were included only once in the analysis.

3.4.4 Ethical consideration

Ethical approval was obtained from the Human Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town (HREC ref: 062/2011) and the Western Cape Provincial Child Health Research committee. Mothers provided written informed consent at enrolment.

3.5 RESULTS

3.5.1 Method Comparison

3.5.1.1 Assessment of nucleic acid quality

Table 7 shows a summary of the nucleic acid extraction and quality assessment of the gDNA from each pneumococcal control strain and confirmed by amplification of 16S rRNA and the pneumococcal *lyt*A specific quantitative PCR (76).

Table 7: Pneumococcal control strain validation

Sample number	Pneumococcal Serotype	gDNA Quantity (ng/µl)	gDNA Quality (260/280 ratio)	*16S* rRNA*	*lyt*A PCR result
1	11A	51.1	1.66	+	+
2	6A	326.3	1.89	+	+
3	7F	67.4	1.80	+	+
4	23A	87.3	1.82	+	+
5	9V	73.6	1.77	+	+
6	15B	125.7	1.97	+	+
7	2	157.1	1.97	+	+
8	15F	157.5	1.68	+	+
9	19F	120.7	1.83	+	+
10	3	112.1	1.77	+	+
11	27	93.7	1.77	+	+
12	9N	445.7	1.95	+	+
13	16F	245.4	1.99	+	+
14	33F	181.8	1.84	+	+
15	4	209.6	1.89	+	+
16	12F	22.8	1.89	+	+
17	19A	229.6	1.78	+	+
18	23B	173.9	1.89	+	+
19	15A	242.8	1.88	+	+
20	31	120.5	1.89	+	+
21	28A	38.2	1.93	+	+
22	35B	122.2	1.84	+	+
23	14	201	1.91	+	+
24	12B	77.5	1.83	+	+
25	8	71.7	1.91	+	+
26	33A	147	1.85	+	+
27	46	259.4	2.04	+	+
28	10A	28.4	1.77	+	+
29	35C	48	1.65	+	+
30	18C	83.4	1.76	+	+
31	10F	101.9	1.65	+	+
32	21	11.1	1.72	+	+

33	22F	29.7	1.99	+	+
34	23F	105.1	1.96	+	+
35	6C	113.1	1.76	+	+
36	1	87.2	1.82	+	+
37	5	112.3	1.86	+	+
38	7C	93.2	1.78	+	+
39	11B	88.7	1.89	+	+
40	24B	107.3	1.74	+	+

+, detected; -, not detected. *16S rRNA PCR was performed as described by Woo *et al* (77).

3.5.1.2 Real-time multiplex PCR

Of the 40 pneumococcal control isolates subjected to rmPCR, 25 isolates yielded a positive signal; 15 isolates failed to yield detectable amplification signal (Figure 10). The serotype identities of these isolates were not included in the rmPCR assay. Of the 25 rmPCR positive isolates, results were 100% (25/25) concordant with Quellung confirmed serotypes (Table 8).

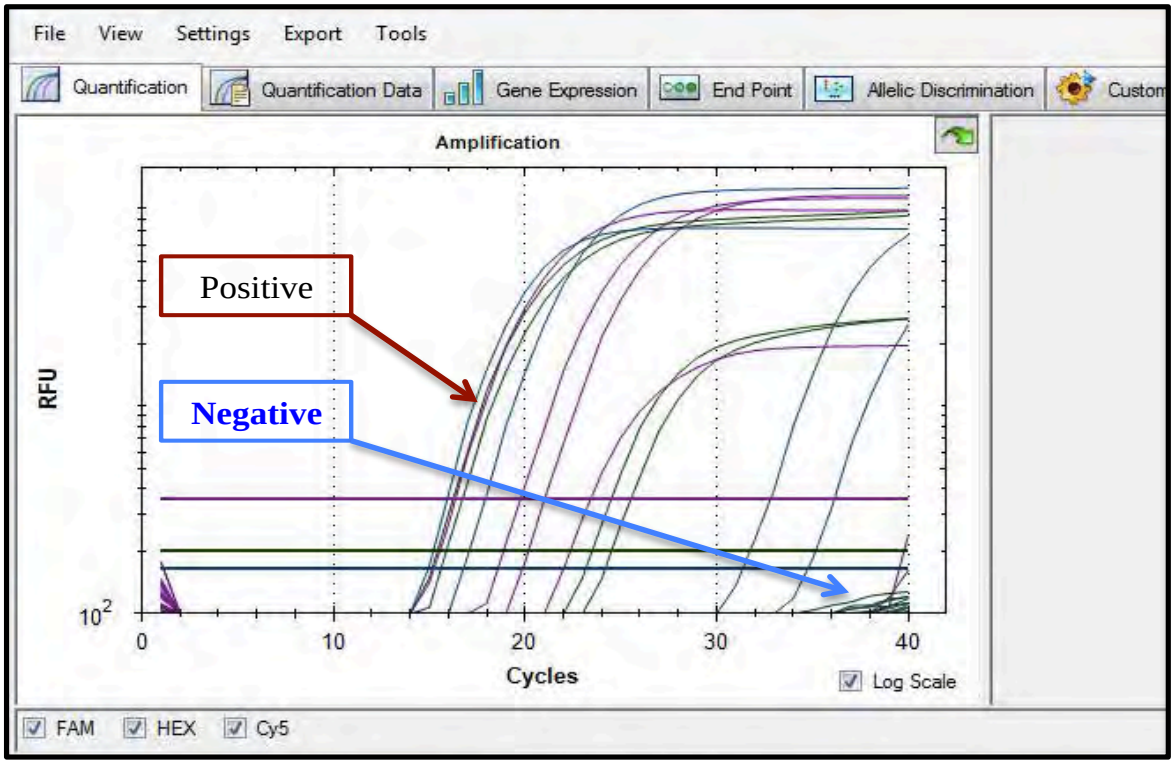

Figure 10: Screenshot of the plate set up and results of the sequential triplex real-time PCR (rmPCR) for molecular typing of pneumococcal control strains. Only pneumococci that were successfully typed had an exponential amplification signals.

Table 8: Concordance of molecular serotyping results of pneumococcal control strains

Serotype	rmPCR[a]	Sequetyping
1	1	1
2	2	No ID[b]
3	3	3
4	4	4
5	5	5
6A	6A/6B	6A
6C	6C/6D	6C/6D
7F	7F/7A	7F/7A
9V	9V/9A	9V
11A	11A	Neg[c]
12B	12F/12A/12B/44/46	12B
12F	12F/12A/12B/44/46	Neg
14	14	14
15A	15A/15F	15A
15F	15A/15F	15F
16F	16F	9V[d]
18C	18C/18A/18B/18F	18B[d]
19A	19A	Neg
19F	19F	19F
22F	22F/22A	Neg
23A	23A	23A
23F	23F	Neg

33A	33F/33A/37	33A/33F/35A
33F	33F/33A/37	33A/33F/35A
46	12F/12A/12B/44/46	12A
7C	N/A*	7C
8	N/A*	8
9N	N/A*	9N
10A	N/A*	10A
10F	N/A*	10C/10F
11B	N/A*	Neg
15B	N/A*	15B/15C
21	N/A*	Neg
23B	N/A*	23B
24B	N/A*	24B
27	N/A*	27
28A	N/A*	28A
31	N/A*	31
35B	N/A*	35B/35C
35C	N/A*	33A/33F/35C

[a]rmPCR: real-time Multiplex PCR. N/A* = serotype not included in the rmPCR panel. No ID[b]. = sequence identity of ≤98% with sequences in GenBank. Neg[c] = negative sequetyping PCR result.

[d]The Sequetyping mistyped serotype18C as 18B, 16F as 9V and 46 as 12A.

3.5.1.3 Sequetyping

Of the 40 pneumococcal control isolates subjected to sequetyping, 33 isolates yielded single amplicons of ~ 1,061 bp; 7 isolates failed to yield detectable amplicons (Figure 11). Sequence analysis yielded 29/33 (88%) sequetype-Quellung concordant results (Table 8). Of the four discordant results, Quellung 16F was identified as 9V by sequetyping, Quellung 46 was identified as 12A by sequetyping while Quellung 18C was identified as 18B. The fourth discordant isolate (Quellung 2) yielded no match (>98%) when submitted to GenBank. The *wzh* PCR was negative in 7/40 (18%) control isolates tested, which consequently could not be sequetyped. Detailed results are provided in Table 8 above.

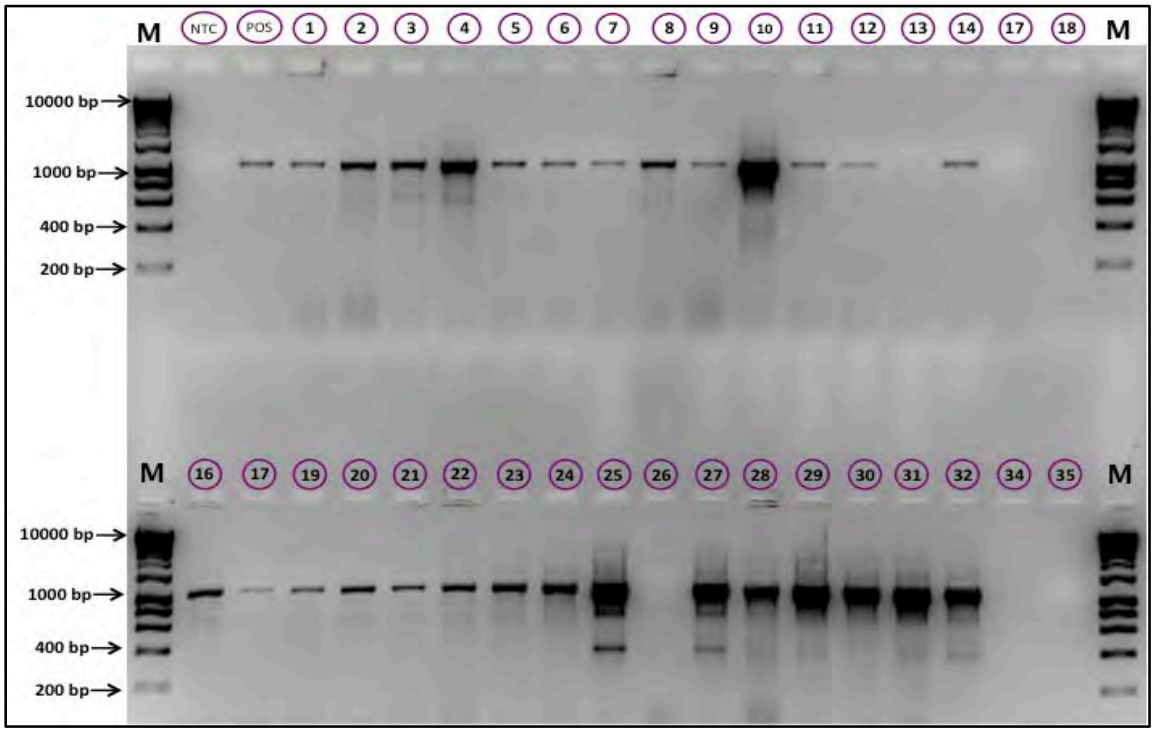

Figure 11: Agarose gel electrophoresis of amplicons generated by the sequetyping assay. The figure shows ethidium bromide-stained amplicons of pneumococcal control strains including controls generated by the sequetyping assay. Lanes 1, 20, 21 and 40: 1-Kb molecular marker (Bioline); Lane 2: Non template control; Lane 3: pneumococcal ATCC46919; Lanes 4-19 and 22-39: pneumococcal control strains.

3.5.1.4 Carriage isolates

Of 135 pneumococcal isolates tested, 132 (98%) were assigned a serotype/-group by the Quellung reaction (Tables 9 and 10). Three (3) isolates could not be typed by either Quellung or molecular methods. A total of 69 (52%) isolates were assigned a serotype covered by the rmPCR assay. Of these, the rmPCR assay assigned the correct serotype to all 69 isolates (Tables 9 and 10).

Table 9: Comparison of serotyping methods with testing performed on carriage isolates

Serotype (n)[a]	rmPCR (n) [b]	Sequetyping (n)	Remarks
1 (1)	100% (1)	100% (1)	
3 (1)	100% (1)	100% (1)	
4 (1)	100% (1)	100% (1)	

6A (2)	100% (2)	50% (1)	2 rmPCR as 6A/6B; 1 sequetyped as 9V
6B (11)	100% (11)	100% (11)	11 rmPCR as 6A/6B
6C (1)	100% (1)	100% (1)	1 rmPCR as 6C/6D; 1 sequetyped as 6C/6D
10A/11A (1)	100% (1)	100% (1)	1 rmPCR as 11A/11D; 1 sequetyped as 10A
11A (8)	100% (8)	75% (6)	8 rmPCR as 11A/11D; 6 sequetyped as 11A/11D/18F; 2 no amplicon in sequetyping
14 (5)	100% (5)	100% (5)	
15A (10)	100% (10)	100% (10)	10 rmPCR as 15A/15F
16F (8)	100% (8)	63% (5)	3 sequetyped as 9V
17F/1¥ (2)	100% (2)	100% (2)	2 rmPCR as 1; 2 sequetyped as 1
18C (3)	100% (3)	33% (1)	3 rmPCR as 18A/18B/18C/18F; 1 sequetyped as 18B, 2 sequetyped with low identity score
19A (8)	100% (8)	75% (6)	2 no amplicon in sequetyping
19F (5)	100% (5)	100% (5)	
22F (2)	100% (2)	100% (2)	2 rmPCR as 22A/22F; 2 sequetyped as 22A/22F
7C (1)*	N/A	100% (1)	
9N (4)*	N/A	100% (4)	
10A (8)*	N/A	100% (8)	
13 (14)*	N/A	93% (13)	1 sequetyped as 15B; 13 sequetyped as 13/20
15B (7)*	N/A	100% (7)	7 sequetyped as 15B/15C
15C (7)*	N/A	100% (7)	7 sequetyped as 15B/15C
17F (3)*	N/A	33% (1)	2 sequetyped as 33C
19B (2)*	N/A	100% (2)	1 sequetyped as 19C, 1 sequetyped as 19F
21 (5)*	N/A	100% (5)	
25A/38 (3)*[c]	N/A	**Neg**[d]	3 no amplicon in sequetyping
35A (9)*	N/A	89% (8)	8 sequetyped as 33A/33F/35A, 1 sequetyped as 13/20
***OMNI NEG* (3)**	*Neg*	*Neg*	
Total (132)	69	112	

The numbers in closed brackets indicate the correct identification of a Quellung-confirmed serotype by the rmPCR and sequetyping assays; [b]rmPCR: real-time multiplex PCR; * = serotypes not included in rmPCR assay. ¥ Mixed serotypes detected; **Neg**[d] = negative sequetyping PCR result; [c]Serotype 25A and 38 were undistinguishable by Quellung and hence reported as 25A/38 (68).

Table 10: Summary of molecular serotyping results of pneumococcal nasopharyngeal carriage isolates compared with the serotype results determined by the Quellung reaction

		rmPCR[a], correctly serotyped[b]			Sequetyping, correctly serotyped[b]			Total
		Yes	No	Negative	Yes	No	Negative	
Quellung	Typable	69	0	63¥	112	13	7	132
	nontypable	0	0	3	0	0	3	3
Total		69	0	66	112	13	10	135

[a] rmPCR: real-time multiplex PCR. [b] isolates typed correctly to the serogroup level compared with phenotypic Quellung reaction results. [c] Negative: no amplification. ¥ all serotypes not covered by the rmPCR panel.

Of the 135 pneumococcal nasopharyngeal isolates that were sequetyped, 125 isolates yielded single amplicons of ~ 1,061bp. A correct serotype/-group was determined in 112 (85%) of the 132 nasopharyngeal isolates. The partial *wzh* sequence of 2/3 Quellung 18C isolates did not match any of the pneumococcal *wzh* sequences in GenBank with >98% identity while the third was determined as 18B. The *wzh* PCR was negative for seven isolates of which three were serotype 25A/38; two were serotype 11A and two were serotype 19A, as confirmed by Quellung testing. Consistent misidentifications by sequetyping, occurring in more than one isolate, were observed for serotype 16F (two isolates sequetyped as 9V) and for serotype 17F (two isolates sequetyped as 33C).

Figure 12 shows the serotype distribution of the pneumococcal isolates, excluding duplicate isolates of the same serotype from the same infant. The most frequently identified serotypes were 11A (9 infants), 13 (8 infants), 15B, 15C (both 7 infants), 16F and 10A (both 6 infants). Of the 91 isolates (the total number of isolates when calculating each serotype only once per child), 22 (24%) were serotypes included in PCV-13 while 69 (76%) serotypes were not.

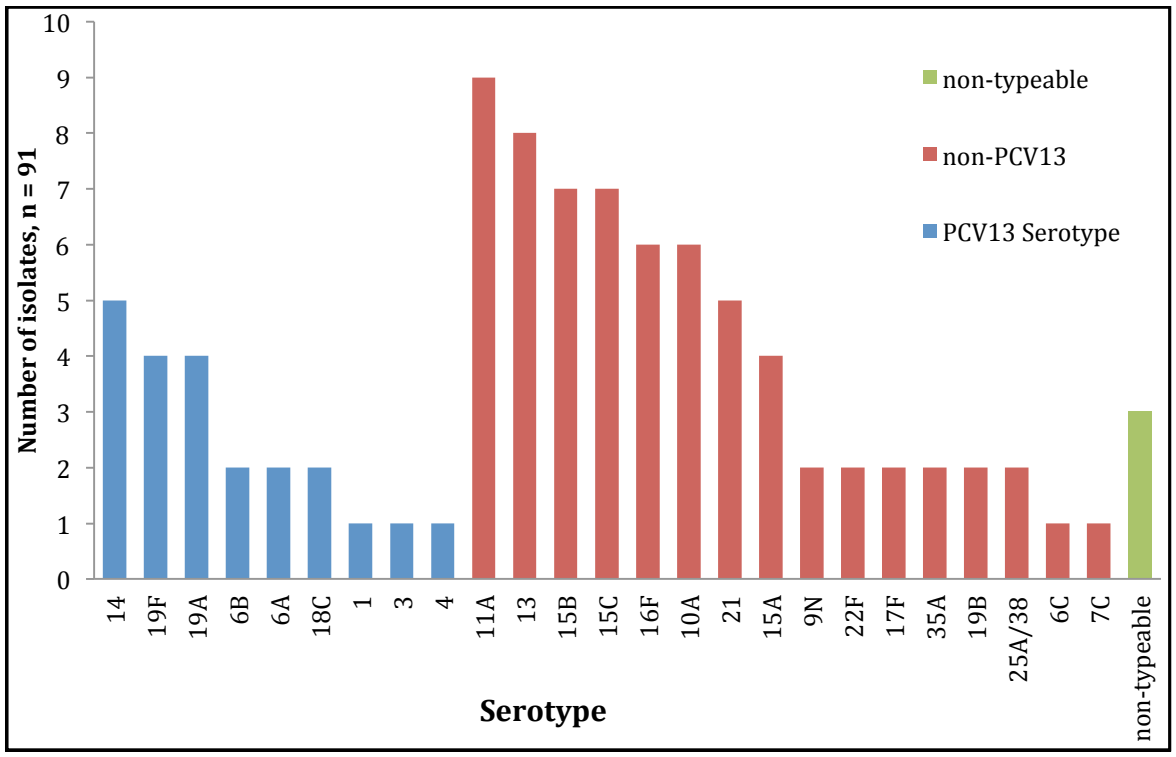

Figure 12: Serotype distribution of nasopharyngeal pneumococcal isolates from the Drakenstein Child Health Study, determined by Quellung reaction. The figure includes all serotypes detected excluding duplicate serotypes from the same infant. Blue = serotypes included in PCV-13; Red= serotypes not included in PCV-13. Green = non-typable isolates.

3.5.2 Discordant Analysis

3.5.2.1 Next-generation sequencing

A total of 14.3 million paired-end sequence reads (2 x 250) were obtained for the four samples as shown in table 11. The quality control steps used preserved the sequence number though reducing the sequence read length to 230 forward, and 120 reverse (230 – 120 fr) respectively.

Table 11: NGS data and assembly metrics

Isolate ID	Paired sequence reads	Number of contigs	N50 (Kb)	Draft genome size (Mb)	Sequencing coverage
9V	1 688 340	52	57.1	2.1	144x
16F	4 621 214	47	68.1	2.1	368x
103347	2 226 252	62	77.3	2.1	184x
103385	5 773 884	56	82.3	2.1	437x

One of pneumococcal control strains and two DCHS strains were serotyped as 16F by Quellung and rmPCR, but mistyped as 9V by sequetyping. A comparison of the *cps* gene loci showed that the *wzh* sequence of all the three queried 16F strains was entirely 9V-like (Figure 13). This is in contrast to the rest of their *cps* loci: which in terms of structural gene organization as well as specific sequence of these genes were entirely 16F-like. Comparative genome analyses of the annotated gene structure showed a marked clustering for the other three queried 16F serotypes all were significantly different from the 9V reference (Figure 14). MLST loci of these 16F strains showed a shared a 16F-like-MLST-type, except the third strain (103385) which had a unique glutamate dehydrogenase gene (*gdh*) allele. The allelic profiles including the corresponding sequence types (ST) are shown in table 12.

Table 12: MLST profile of discordant serotype 16F strains

Sample	ST	*aroE**	*ddl*	*gdh*	*gki*	*recP*	*spi*	*xpt*
9V_NICD	6530	1	6	11	10	5	6	1
16F_NICD	5326	50	31	9	15	1	36	3
103347	4088	50	14	8	1	16	6	88
103385	??	50	31	??	15	1	36	3

*The estimated MLST allele sizes are *aro*E = 405bp, ddl = 441bp, gdh = 460bp, *gki* = 483bp, recP = 450bp, spi = 474bp and *Xpt* = 486bp.?**? =** Not determined.

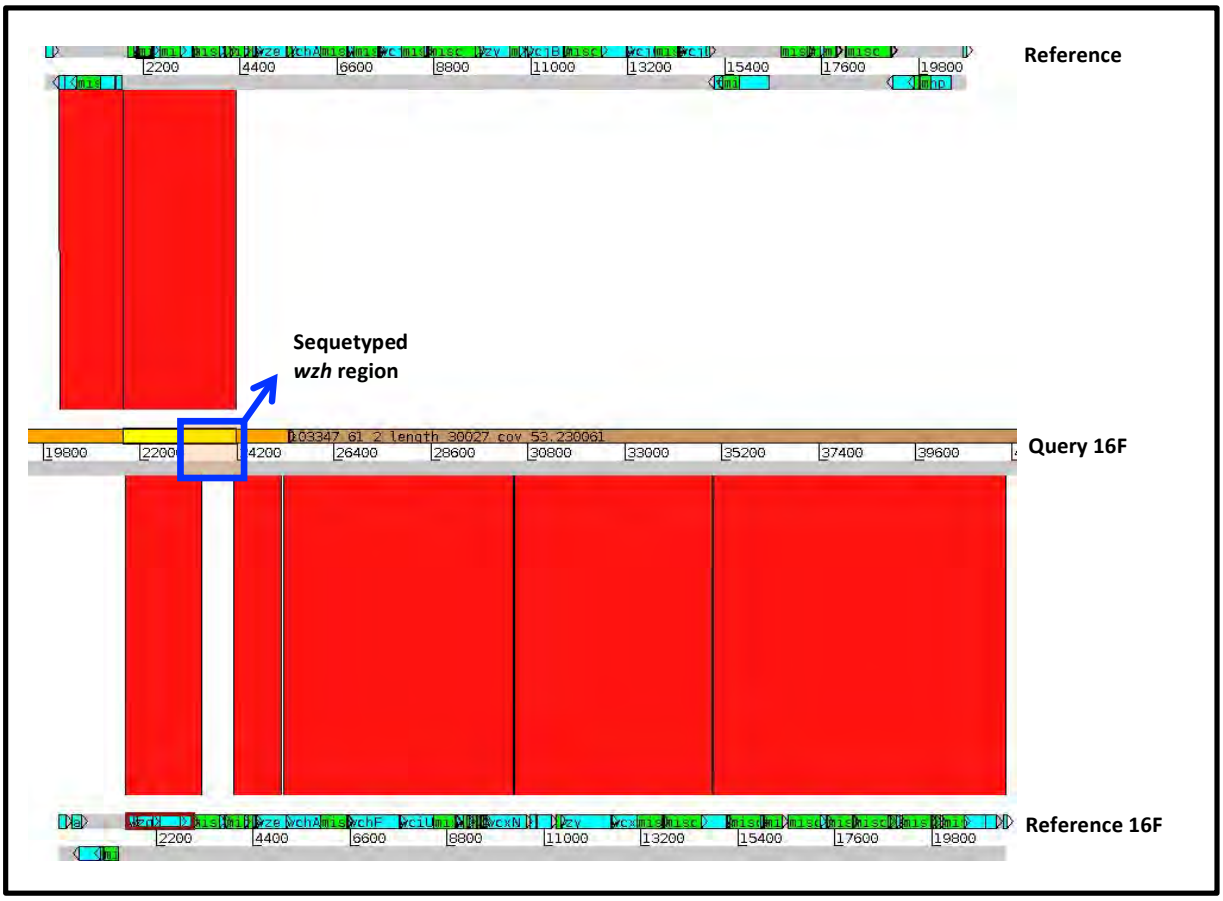

Figure 13: Similarity of 16F-like capsular polysaccharide (*cps*) gene loci. Sequences from pneumococci serotyped as 16F Quellung but sequetyped as 9V were compared to reference 9V (CR931648) and 16F (CR931668) *cps* sequences. Artemis Comparison Tool (ACT) was used to generate and view gene homology. The top lines represent the forward and reverse strand of a serotype 9V reference, the middle lines represent the queried 16F strain and the bottom lines shows the 16F reference. The portion of the *wzh* gene that is amplified by the sequetyping assay is shown by the blue rectangle. The clear blocks below the blue box shows regions where the genes are not similar. BLASTN matches are shown as red bands between sequences, indicating the degree of similarity between the sequences.

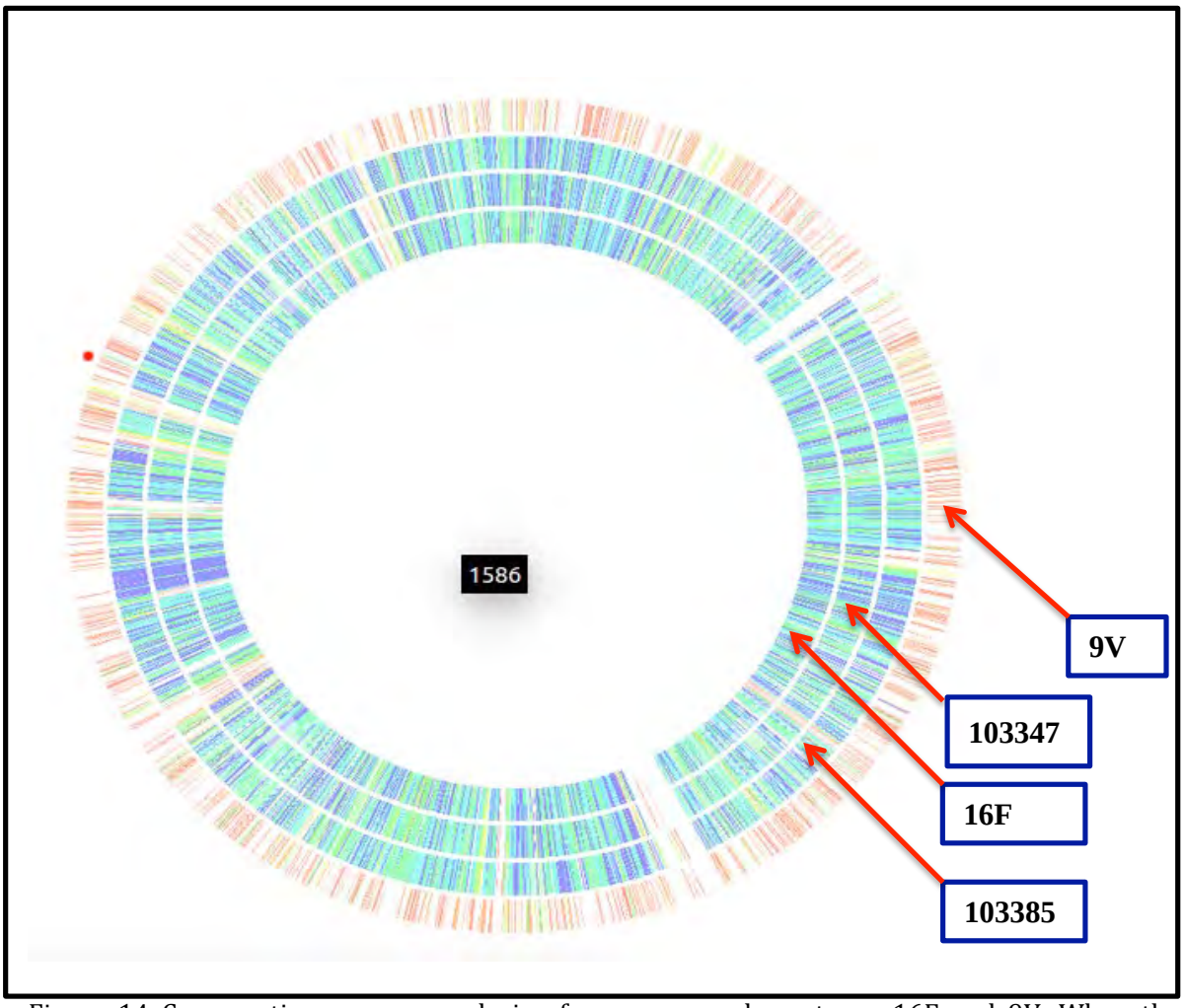

Figure 14 Comparative genome analysis of pneumococcal serotypes 16F and 9V. When the sequence identities of all four genomes were compared using RAST (Rapid Annotation using Subsystem Technology), the genome backbone of all three 16F (103347 and 103385 from this study and a 16F control strain) were mostly identical but divergent from 9V. The colour codes represent how close or divergent the genomes are. Therefore, similar genome backgrounds will have similar colours.

3.6 DISCUSION

To identify the accuracy of a rapid high throughput molecular serotyping assay, rmPCR was compared to sequetyping using first a panel of 40 control isolates and then 132 carriage isolates.

rmPCR:

The rmPCR which we used is designed to detect and identify 21 serotypes including all serotypes/-groups in PCV-13, all of which were included in our analysis. Concordance with Quellung was 100% (25/25) for those control isolates included in the rmPCR panel.

Amongst the pneumococcal carriage isolates tested, the correct serotype/-group could be assigned to 52% of typable isolates (69/132) by rmPCR. 47% (63/132) of isolates were not included in the rmPCR panel, and therefore the rmPCR was 100% accurate for those isolates included in the panel. The high number of negative results from rmPCR amongst nasopharyngeal isolates was not surprising since this assay is likely to be less useful in areas where pneumococcal conjugate vaccines have been implemented resulting in serotype replacement which may arise as a result of either serotype unmasking or capsular switching. Data from the United States on invasive pneumococcal isolates showed a decline in serotypes included in the rmPCR assay from 92% (3812/4106) prior to PCV-7 implementation to 79% (2939/3708) after PCV-7 roll out and a further decrease to 74% (2581/3480) post PCV-13 implementation (Unpublished US Active Bacterial Core surveillance data). Amongst our small cohort, non-vaccine serotypes 11A, 13, 15B, 15C, 16F and 10A were the most prevalent serotypes identified. Similarly, data from a number of other post PCV-13 surveillance studies have reported serotypes 11A, 15A/B/C, 16, but also 22F, 21 and 34 as prevalent

non-vaccine serotypes (78–85). The original rmPCR protocol (39) did not make reference to any internal control in the assay set up. However, as part of our assay set-up and validation, we screened all the samples for 16S rRNA PCR to check for inhibition (77) and subjected all rmPCR negative samples to *cps*A (*wzg*) PCR to check the integrity of the capsulation locus (41). Although relatively uncommon, 1-2%, *cps*A-negative serotypeable pneumococci have been shown to occur especially in serotype 25A/F, 38 and to some extent 14 and 35A (86). The *cps*A primer sequences can not detect serotype 25A/F and 38 due to mutations or 5' gene rearrangement which is characterised by the insertion of transposase genes between *cps*CD and *cps*AB (87).

Sequetyping:

Sequetyping is designed to identify up to 46 different serotypes/-groups; concordance with Quellung for the 40 control strains was 88% (29/33), with failure of sequetyping PCR for 7 strains and misidentification of 4 serotypes. Amongst nasopharyngeal isolates, 85% of typable isolates (112/132) were correctly identified by sequetyping. For 7 isolates no PCR product was obtained and sequetyping misidentified 13 isolates. Based on published sequence data one would have expected the sequetyping assay to amplify and classify all except three nasopharyngeal isolates (Serotypes 25A and 38) into the correct serotype/-group with some strains giving ambiguous results (serotype 13/20, 17F/33C, and 11A/D/1818F).

The broad range of serotypes that are theoretically detectable by sequetyping is a major advantage of this technique. It is not clear why, in this study, amplification failed for 7/40 (18%) control strains. Interrogation of published gene sequences for these serotypes indicated that these serotypes should generate PCR products with the protocol used here (38). PCR inhibition was excluded based on successful *lyt*A PCR in all 7 strains. Interestingly, four of the nasopharyngeal carriage isolates that failed to

amplify during sequetyping were serotypes for which a similar problem was encountered when sequetyping the control isolates (serotypes 11A and 19A). The remaining three sequetype-negative isolates were of serotype 25A/38 which were expected to be non-amplifiable because of absence of the reverse primer binding site in the *wzd* gene (1,38,87). Sequetyping misidentified three control strains (for which Quellung and rmPCR were concordant). The sequence obtained from *wzh* amplicon of serotype 2 did not match any of the *wzh* nucleotide sequences in GenBank with >98% sequence identity. The original study describing sequetyping did not test this serotype although their *insilico* analysis had predicted that the primer sets should be able to amplify serotype 2 (38). Further tests are needed to resolve this, particularly re-sequencing the entire capsulation locus of this serotype. Misidentification of the serotype 46 isolate as 12A is explained by high relatedness between these serotypes as their *cps* gene clusters are almost identical (1). Based on our observation of mistyping the 18C PCV-13 serotype as 18B by sequetyping, it may be warranted to confirm all 18B results by Quellung. Misidentification of 17F isolates as 33C was predicted by the original sequetyping paper as these serotypes cannot be distinguished based on their *wzh* sequence (38).

We found one control strain and three nasopharyngeal strains that were serotyped as 16F by Quellung, but sequetyped as 9V. Even though the *wzh* sequences of the queried 16F was entirely 9V-like, the serotype specific *wzy/wzx* genes were entirely 16F-like. Based on analysis of the core genome, the 16F control strain was identified as sequence type (ST) 5326, one of the nasopharyngeal isolates was identified as ST4088, while the other nasopharyngeal isolates was a new ST, which was a single-locus variant of ST5326. These sequence types are all commonly associated with serotype 16F (75). Therefore, our strains seem to be 16F strains in almost every sense; they only have a

9V-*wzh* gene. This structural difference is not expected to have occurred as in response to vaccine pressure, because none of the currently used vaccine formulations include serotype 16F and the exchanged 9V gene does not result in a modified phenotype. In practice, in our setting, each isolate with a 9V sequetype result should be investigated further.

Both molecular assays are able to type many pneumococcal strains only to the serogroup level. Discrimination of individual serotypes within a serogroup may be important for more detailed assessment of carriage and vaccine effectiveness. When selecting a serotyping method, test characteristics other than accuracy may also be relevant. The sequetyping assay, which involves a single amplification step, is inexpensive compared with the rmPCR assay, which is labour intensive, includes many costly PCR probes and is constrained by the limited multiplexing options of real-time PCR. Interpretation of the sequetyping results is based on the publically available GenBank database. An advantage of this is its free accessibility, but the uncontrolled and changing nature of this database could be a risk for the assignment of serotypes. Our automated 'sequetyper' application makes analysis of the relevant sequence data for sequetyping rapid and simple. A significant disadvantage of sequetyping is that the targeted *wzh* gene is not serotype-specific and does itself not determine serotype – the results are inferred based on association. It is entirely feasible therefore (as we found here) that for specific serotypes and in particular populations of pneumococci that this association may not correctly predict serotype. The technique is therefore likely only useful for typing pneumococci from populations of pneumococci where such association has been confirmed using another typing technique. In our case this would mean confirming serotype for a smaller subset such as serotype 9V, 13, 20 and serogroup 33. The CDC Streptococcal laboratory has recently provided an extended version of the

conventional multiplex PCR assay (not available at the time of this study) that utilises 41 serotype-specific primer sets to detect up to 70 different pneumococcal serotypes (http://www.cdc.gov/streplab/downloads/pcr-oligonucleotide-primers.pdf). The basis for this assay is similar to the rmPCR employed here although less costly. The benefits of deducing more than 70 serotypes by this assay needs to be weighed against the fairly labour intensive protocol and risk of amplicon contamination. A comprehensive multicenter study has recently compared the 20 pneumococcal serotyping methods and shown that micro-arrays (with a culture amplification step) had the highest sensitivity and positive predictive value (95.8% and 93.7% respectively) compared to other methods (6). In addition, microarray performed better at detecting multiple pneumococcal serotypes including the less predominant serotypes in instances where there were multiple serotypes.

In conclusion, sequetyping is a useful technique for large scale molecular serotyping of pneumococcal strains, particularly post-PCV introduction, because of the broad range of non-vaccine serotypes that can be detected, low cost and ease of use. Our results suggest the need for an extended and carefully curated database of serotype-specific sequence data, which will increase the accuracy and expand the serotype coverage of the sequetyping method. However, given the potential for gene exchange that could result in false assignment of serotype by sequetyping, it is necessary to confirm serotype assignment using a different method. This may still be cost-saving as it would involve testing only the specific serotype assigned by serotyping, using the Quellung method, for example. The rmPCR assay, ideally extended to include more serotypes is reliable but cost, time required to perform testing, and currently restricted serotype coverage may limit its widespread application for large epidemiological studies. In the future it is likely that WGS will be increasingly used as a tool for serotype inference.

WGS has many advantages, in that additional information (such as multi-locus sequence type and antimicrobial resistance) can be inferred from the same dataset without additional testing, and that serotype can be definitively assigned. As sequence costs decline further, bioinformatics pipelines are increasingly automated and the technology is more widely available in low-resource settings it is likely that WGS will replace conventional typing tools for pneumococci.

3.7 REFERENCES

1. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. 2006;2(3):0262–9.
2. Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, et al. Biochemical, Genetic, and Serological Characterization of Two Capsule Subtypes among *Streptococcus pneumoniae* Serotype 20 Strains: Discovery Of A New Pneumococcal Serotype. J Biol Chem [Internet]. 2012 Jun 26 [cited 2015 Jun 25];287(33):27885–94.
3. Bogaert D, De Groot R, Hermans PWM. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis [Internet]. 2004 Mar [cited 2014 May 1];4(3):144–54.
4. Coles CL, Sherchand JB, Khatry SK, Katz J, Leclerq SC, Mullany LC, et al. Nasopharyngeal carriage of S. pneumoniae among young children in rural Nepal. Trop Med Int Health [Internet]. 2009 Sep [cited 2014 May 11];14(9):1025–33.
5. Antonio M, Dada-Adegbola H, Biney E, Awine T, O'Callaghan J, Pfluger V, et al. Molecular epidemiology of pneumococci obtained from Gambian children aged 2-29 months with invasive pneumococcal disease during a trial of a 9-valent pneumococcal conjugate vaccine. BMC Infect Dis [Internet]. 2008 Jan [cited 2014 May 11];8:81.
6. Brugger SD, Frey P, Aebi S, Hinds J, Mühlemann K. Multiple colonization with S. pneumoniae before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. PLoS One [Internet]. 2010 Jan [cited 2014 May 11];5(7):e11638.
7. Rivera-Olivero IA, Blommaart M, Bogaert D, Hermans PWM, de Waard JH. Multiplex PCR reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuela. J Med Microbiol [Internet]. 2009 May [cited 2014 May 10];58(Pt 5):584–7.
8. Turner P, Hinds J, Turner C, Jankhot A, Gould K, Bentley SD, et al. Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. J Clin Microbiol [Internet]. 2011 May [cited 2014 May 11];49(5):1784–9.
9. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science [Internet]. 2011 Jan 28 [cited 2015 Jul 15];331(6016):430–4.
10. Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, et al. Selective and genetic constraints on pneumococcal serotype switching. PLoS Genet [Internet]. 2015 Feb [cited 2015 Apr 7];11(3):e1005095.
11. Croucher NJ, Chewapreecha C, Hanage WP, Harris SR, McGee L, van der Linden M, et al. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. Genome Biol Evol [Internet]. 2014 Jul [cited 2015 May 20];6(7):1589–602.
12. Austrian R. The quellung reaction, a neglected microbiologic technique. Mt Sinai J Med [Internet]. 1960 [cited 2014 May 12];43(6):699–709.
13. Sørensen UB. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol [Internet]. 1993 Aug [cited 2015 Jan 23];31(8):2097–100.

14. Henrichsen J. Six newly recognized types of *Streptococcus pneumoniae*. J Clin Microbiol [Internet]. 1995 Oct [cited 2015 Jan 23];33(10):2759–62.
15. Satzke C, Turner P, Virolainen-Julkunen A, Adrian P V., Antonio M, Hare KM, et al. Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. Vaccine [Internet]. 2013 Dec 17 [cited 2014 May 3];32(1):165–79.
16. Porter BD, Ortika BD, Satzke C. Capsular Serotyping of *Streptococcus pneumoniae* by latex agglutination. J Vis Exp [Internet]. 2014 Jan [cited 2015 Aug 14];(91):51747.
17. Browne K, Miegel J, Stottmeier KD. Detection of pneumococci in blood cultures by latex agglutination. J Clin Microbiol [Internet]. 1984 May [cited 2015 Aug 17];19(5):649–50.
18. Slotved H-C, Kaltoft M, Skovsted IC, Kerrn MB, Espersen F. Simple, rapid latex agglutination test for serotyping of pneumococci (Pneumotest-Latex). J Clin Microbiol [Internet]. 2004 Jun [cited 2015 Aug 17];42(6):2518–22.
19. Fenoll A, Jado I, Vicioso D, Casal J. Dot blot assay for the serotyping of pneumococci. J Clin Microbiol [Internet]. 1997 Mar [cited 2014 May 11];35(3):764–6.
20. Bogaert D, Veenhoven RH, Sluijter M, Sanders EAM, de Groot R, Hermans PWM. Colony blot assay: a useful method to detect multiple pneumococcal serotypes within clinical specimens. FEMS Immunol Med Microbiol [Internet]. 2004 Jul 1 [cited 2015 Aug 17];41(3):259–64.
21. Bronsdon MA, O'Brien KL, Facklam RR, Whitney CG, Schwartz B, Carlone GM. Immunoblot method to detect *Streptococcus pneumoniae* and identify multiple serotypes from nasopharyngeal secretions. J Clin Microbiol [Internet]. 2004 Apr [cited 2014 May 11];42(4):1596–600.
22. Lankinen KS, Rintamäki S, Syrjänen R, Kilpi T, Ruutu P, Leinonen M. Type-specific enzyme immunoassay for detection of pneumococcal capsular polysaccharide antigens in nasopharyngeal specimens. J Microbiol Methods [Internet]. 2004 Feb [cited 2015 Aug 17];56(2):193–9.
23. Leeming JP, Cartwright K, Morris R, Martin SA, Smith MD. Diagnosis of invasive pneumococcal infection by serotype-specific urinary antigen detection. J Clin Microbiol [Internet]. 2005 Oct [cited 2015 Aug 17];43(10):4972–6.
24. 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 [Internet]. 1991 May [cited 2015 Aug 17];163(5):1094–102.
25. Park MK, Briles DE, Nahm MH. A latex bead-based flow cytometric immunoassay capable of simultaneous typing of multiple pneumococcal serotypes (Multibead assay). Clin Diagn Lab Immunol [Internet]. 2000 May [cited 2015 Aug 17];7(3):486–9.
26. Yu J, Lin J, Benjamin WH, Waites KB, Lee C, Nahm MH. Rapid multiplex assay for serotyping pneumococci with monoclonal and polyclonal antibodies. J Clin Microbiol [Internet]. 2005 Jan [cited 2015 Jan 23];43(1):156–62.
27. Lin J, Kaltoft MS, Brandao AP, Echaniz-Aviles G, Brandileone MCC, Hollingshead SK, et al. Validation of a multiplex pneumococcal serotyping assay with clinical samples. J Clin Microbiol [Internet]. 2006 Feb [cited 2015 Jan 23];44(2):383–8.
28. Yu J, Lin J, Kim K-H, Benjamin WH, Nahm MH. Development of an automated and

multiplexed serotyping assay for *Streptococcus pneumoniae*. Clin Vaccine Immunol [Internet]. 2011 Nov [cited 2015 Jan 23];18(11):1900–7.
29. Fossieck B, Craig R, Paterson PY. Counterimmunoelectrophoresis for rapid diagnosis of meningitis due to Diplococcus pneumoniae. J Infect Dis [Internet]. 1973 Jan [cited 2015 Aug 17];127(1):106–9.
30. Russell H, Facklam RR, Padula JF, Cooksey R. Capillary precipitin typing of *Streptococcus pneumoniae.* J Clin Microbiol [Internet]. 1978 Oct [cited 2015 Aug 17];8(4):355–9.
31. Wicher K, Kalinka C, Mlodozeniec P, Rose NR. Fluorescent antibody technic used for identification and typing of *Streptococcus pneumoniae.* Am J Clin Pathol [Internet]. 1982 Jan [cited 2015 Aug 17];77(1):72–7.
32. Billal DS, Hotomi M, Suzumoto M, Yamauchi K, Arai J, Katsurahara T, et al. Determination of pneumococcal serotypes/genotypes in nasopharyngeal secretions of otitis media children by multiplex PCR. Eur J Pediatr [Internet]. 2008 Apr [cited 2014 May 11];167(4):401–7.
33. Donkor ES, Bishop CJ, Gould KA, Hinds J, Antonio M, Wren B, et al. High levels of recombination among *Streptococcus pneumoniae* isolates from the Gambia. MBio [Internet]. 2011 Jan [cited 2014 May 11];2(3):e00040–11.
34. Elberse KEM, van de Pol I, Witteveen S, van der Heide HGJ, Schot CS, van Dijk A, et al. Population structure of invasive *Streptococcus pneumoniae* in The Netherlands in the pre-vaccination era assessed by MLVA and capsular sequence typing. PLoS One [Internet]. 2011 Jan [cited 2014 May 11];6(5):e20390.
35. Kong F, Gilbert GL. Using cpsA-cpsB sequence polymorphisms and serotype-/group-specific PCR to predict 51 *Streptococcus pneumoniae* capsular serotypes. J Med Microbiol [Internet]. 2003 Dec [cited 2014 May 11];52(Pt 12):1047–58.
36. Kong F, Wang W, Tao J, Wang L, Wang Q, Sabananthan A, et al. A molecular-capsular-type prediction system for 90 *Streptococcus pneumoniae* serotypes using partial cpsA-cpsB sequencing and wzy- or wzx-specific PCR. J Med Microbiol [Internet]. 2005 Apr [cited 2014 May 11];54(Pt 4):351–6.
37. Kong F, Brown M, Sabananthan A, Zeng X, Gilbert GL. Multiplex PCR-based reverse line blot hybridization assay to identify 23 *Streptococcus pneumoniae* polysaccharide vaccine serotypes. J Clin Microbiol [Internet]. 2006 May [cited 2015 Jan 23];44(5):1887–91.
38. Leung MH, Bryson K, Freystatter K, Pichon B, Edwards G, Charalambous BM, et al. Sequetyping: serotyping *Streptococcus pneumoniae* by a single PCR sequencing strategy. J Clin Microbiol [Internet]. 2012 Jul [cited 2014 May 11];50(7):2419–27.
39. Pimenta FC, Roundtree A, Soysal A, Bakir M, du Plessis M, Wolter N, et al. Sequential triplex real-time PCR assay for detecting 21 pneumococcal capsular serotypes that account for a high global disease burden. J Clin Microbiol [Internet]. 2013 Feb [cited 2014 May 11];51(2):647–52.
40. Brito DA, Ramirez M, de Lencastre H. Serotyping *Streptococcus pneumoniae* by multiplex PCR. J Clin Microbiol [Internet]. 2003 Jun [cited 2014 May 11];41(6):2378–84.
41. Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. J Clin Microbiol [Internet]. 2006 Jan [cited 2015 Jan 22];44(1):124–31.
42. Batt SL, Charalambous BM, McHugh TD, Martin S, Gillespie SH. Novel PCR-restriction fragment length polymorphism method for determining serotypes or serogroups of *Streptococcus pneumoniae* isolates. J Clin Microbiol [Internet]. 2005

Jun [cited 2015 Jun 26];43(6):2656–61.
43. Selva L, del Amo E, Brotons P, Muñoz-Almagro C. Rapid and easy identification of capsular serotypes of *Streptococcus pneumoniae* by use of fragment analysis by automated fluorescence-based capillary electrophoresis. J Clin Microbiol [Internet]. 2012 Nov [cited 2015 Jan 23];50(11):3451–7.
44. Massire C, Gertz RE, Svoboda P, Levert K, Reed MS, Pohl J, et al. Concurrent serotyping and genotyping of pneumococci by use of PCR and electrospray ionization mass spectrometry. J Clin Microbiol [Internet]. 2012 Jun [cited 2015 Jan 23];50(6):2018–25.
45. Zhou F, Kong F, Tong Z, Gilbert GL. Identification of less-common *Streptococcus pneumoniae* serotypes by a multiplex PCR-based reverse line blot hybridization assay. J Clin Microbiol [Internet]. 2007 Oct [cited 2015 Aug 15];45(10):3411–5.
46. O'Sullivan MVN, Zhou F, Sintchenko V, Kong F, Gilbert GL. Multiplex PCR and reverse line blot hybridization assay (mPCR/RLB). J Vis Exp [Internet]. 2011 Jan [cited 2015 Jun 21];(54).
47. Tarragó D, Fenoll A, Sánchez-Tatay D, Arroyo LA, Muñoz-Almagro C, Esteva C, et al. Identification of pneumococcal serotypes from culture-negative clinical specimens by novel real-time PCR. Clin Microbiol Infect [Internet]. 2008 Sep [cited 2015 Aug 15];14(9):828–34.
48. Azzari C, Moriondo M, Indolfi G, Cortimiglia M, Canessa C, Becciolini L, et al. Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. PLoS One [Internet]. 2010 Jan [cited 2014 May 11];5(2):e9282.
49. Moore CE, Sengduangphachanh A, Thaojaikong T, Sirisouk J, Foster D, Phetsouvanh R, et al. Enhanced determination of *Streptococcus pneumoniae* serotypes associated with invasive disease in Laos by using a real-time polymerase chain reaction serotyping assay with cerebrospinal fluid. Am J Trop Med Hyg [Internet]. 2010 Sep [cited 2014 May 11];83(3):451–7.
50. Yildirim I, Hanage WP, Lipsitch M, Shea KM, Stevenson A, Finkelstein J, et al. Serotype specific invasive capacity and persistent reduction in invasive pneumococcal disease. Vaccine [Internet]. 2010 Dec 16 [cited 2015 Jan 23];29(2):283–8.
51. Blaschke AJ. Interpreting assays for the detection of *Streptococcus pneumoniae.* Clin Infect Dis [Internet]. 2011 May [cited 2014 May 11];52 Suppl 4:S331–7.
52. Resti M, Moriondo M, Cortimiglia M, Indolfi G, Canessa C, Becciolini L, et al. Community-acquired bacteremic pneumococcal pneumonia in children: diagnosis and serotyping by real-time polymerase chain reaction using blood samples. Clin Infect Dis [Internet]. 2010 Nov 1 [cited 2014 May 11];51(9):1042–9.
53. Dhoubhadel BG, Yasunami M, Yoshida L-M, Thi HAN, Thi THV, Thi TAN, et al. A novel high-throughput method for molecular serotyping and serotype-specific quantification of *Streptococcus pneumoniae* using a nanofluidic real-time PCR system. J Med Microbiol [Internet]. 2014 Apr [cited 2015 Jun 26];63(Pt 4):528–39.
54. Tomita Y, Okamoto A, Yamada K, Yagi T, Hasegawa Y, Ohta M. A new microarray system to detect *Streptococcus pneumoniae* serotypes. J Biomed Biotechnol [Internet]. 2011 Jan [cited 2015 Aug 17];2011:352736.
55. Wang Q, Wang M, Kong F, Gilbert GL, Cao B, Wang L, et al. Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes. J

Microbiol Methods [Internet]. 2007 Jan [cited 2015 Aug 17];68(1):128–36.
56. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, et al. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res [Internet]. 2012 Jan [cited 2015 Jul 16];40(1):e6.
57. Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, et al. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*Spain23F ST81. J Bacteriol [Internet]. 2009 Mar [cited 2015 Jul 9];191(5):1480–9.
58. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet [Internet]. 2014 Mar [cited 2015 Aug 10];46(3):305–9.
59. Liyanapathirana V, Ang I, Tsang D, Fung K, Ng TK, Zhou H, et al. Application of a target enrichment-based next-generation sequencing protocol for identification and sequence-based prediction of pneumococcal serotypes. BMC Microbiol [Internet]. 2014 Jan [cited 2015 Aug 17];14:60.
60. Jauneikaite E, Tocheva AS, Jefferies JMC, Gladstone RA, Faust SN, Christodoulides M, et al. Current methods for capsular typing of *Streptococcus pneumoniae.* J Microbiol Methods [Internet]. 2015 Jun [cited 2015 Jun 26];113:41–9.
61. von Gottberg A, Cohen C, de Gouveia L, Meiring S, Quan V, Whitelaw A, et al. Epidemiology of invasive pneumococcal disease in the pre-conjugate vaccine era: South Africa, 2003-2008. Vaccine [Internet]. 2013 Aug 28 [cited 2015 Jan 29];31(38):4200–8.
62. Wasas AD, Huebner RE, De Blanche M, Klugman KP. Long-term survival of *Streptococcus pneumoniae* at room temperature on Dorset egg medium. J Clin Microbiol [Internet]. 1998 Apr [cited 2014 May 12];36(4):1139–40.
63. le Roux DM, Myer L, Nicol MP, Zar HJ. Incidence and severity of childhood pneumonia in the first year of life in a South African birth cohort: the Drakenstein Child Health Study. Lancet Glob Heal [Internet]. 2015 Feb [cited 2015 Jan 28];3(2):e95–103.
64. Dube FS, Kaba M, Whittaker E, Zar HJ, Nicol MP. Detection of *Streptococcus pneumoniae* from Different Types of Nasopharyngeal Swabs in Children. PLoS One [Internet]. 2013 Jan [cited 2014 May 11];8(6):e68097.
65. Winn, W. C., Allen, S. D., Janda, W. M., Koneman, E. W., Procop, G. W., Schreckenberger, P. C., and Woods GL. Gram-Positive Cocci Part II: *Streptococci, Enterococci*, and the "*Streptococcus-lik*e" Bacteria, p. 672-764. In: Konemans's Color Atlas and Textbook of Diagnostic Microbiology. 6th ed. Wilkins LW and, editor. 672-764 p.
66. Leung MHY, Oriyo NM, Gillespie SH, Charalambous BM. The adaptive potential during nasopharyngeal colonisation of *Streptococcus pneumoniae*. Infect Genet Evol [Internet]. 2011 Dec [cited 2015 Jan 16];11(8):1989–95.
67. Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MCC, Nahm MH. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. J Clin Microbiol [Internet]. 2007 Apr [cited 2015 Feb 2];45(4):1225–33.
68. Nunes MC, Jones SA, Groome MJ, Kuwanda L, Van Niekerk N, von Gottberg A, et al. Acquisition of *Streptococcus pneumoniae* in South African children vaccinated with 7-valent pneumococcal conjugate vaccine at 6, 14 and 40 weeks of age. Vaccine [Internet]. 2014 Dec 23 [cited 2015 Jan 2]
69. Andrews S. FastQC a quality-control tool for high-throughput sequence data.

[Internet]. 2011.
70. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics [Internet]. 2014 Apr 28 [cited 2014 Jul 9];30(15):2114–20.
71. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One [Internet]. 2012 Jan [cited 2014 Jul 10];7(2):e30619.
72. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol [Internet]. 2013 Oct [cited 2015 Jan 11];20(10):714–37.
73. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics [Internet]. 2008 Jan [cited 2014 Jul 9];9:75.
74. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 Aug 15 [cited 2014 Jul 9];25(16):2078–9.
75. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics [Internet]. 2010 Jan [cited 2014 Dec 7];11:595.
76. Greve T, Møller JK. Accuracy of using the lytA gene to distinguish *Streptococcus pneumoniae* from related species. J Med Microbiol. 2012 Apr;61(Pt 4):478–82.
77. Woo PC, Cheung EY, Leung K, Yuen K. Identification by 16S ribosomal RNA gene sequencing of an Enterobacteriaceae species with ambiguous biochemical profile from a renal transplant recipient. Diagn Microbiol Infect Dis [Internet]. 2001 Feb [cited 2015 Jan 23];39(2):85–93.
78. Koen N, Wyatt GE, Williams JK, Zhang M, Myer L, Zar HJ, et al. Intimate partner violence: associations with low infant birthweight in a South African birth cohort. Metab Brain Dis [Internet]. 2014 Apr 12 [cited 2014 May 12];
79. Guevara M, Ezpeleta C, Gil-Setas A, Torroba L, Beristain X, Aguinaga A, et al. Reduced incidence of invasive pneumococcal disease after introduction of the 13-valent conjugate vaccine in Navarre, Spain, 2001-2013. Vaccine [Internet]. 2014 May 7 [cited 2014 Apr 28];32(22):2553–62.
80. Olarte L, Hulten KG, Lamberth L, Mason EO, Kaplan SL. Impact of the 13-Valent Pneumococcal Conjugate Vaccine on Chronic Sinusitis Associated with *Streptococcus pneumoniae* in Children. Pediatr Infect Dis J [Internet]. 2014 Apr 25 [cited 2014 May 6];
81. Kaplan SL, Barson WJ, Lin PL, Romero JR, Bradley JS, Tan TQ, et al. Early trends for invasive pneumococcal infections in children after the introduction of the 13-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J [Internet]. 2013 Mar [cited 2014 May 12];32(3):203–7.
82. Steens A, Bergsaker MAR, Aaberge IS, Rønning K, Vestrheim DF. Prompt effect of replacing the 7-valent pneumococcal conjugate vaccine with the 13-valent vaccine on the epidemiology of invasive pneumococcal disease in Norway. Vaccine [Internet]. 2013 Dec 16 [cited 2014 Apr 30];31(52):6232–8.
83. van Hoek AJ, Sheppard CL, Andrews NJ, Waight PA, Slack MPE, Harrison TG, et al. Pneumococcal carriage in children and adults two years after introduction of the thirteen valent pneumococcal conjugate vaccine in England. Vaccine [Internet]. 2014 Mar 21 [cited 2014 May 7];
84. Aguiar S, Brito M, Horacio A, Lopes J, Ramirez M, Melo-Cristino J. Decreasing incidence and changes in serotype distribution of invasive pneumococcal disease

in persons aged under 18 years since introduction of 10-valent and 13-valent conjugate vaccines in Portugal, July 2008 to June 2012. Euro Surveill [Internet]. 2014 Jan [cited 2014 May 12];19(12).
85. Cohen R, Levy C, Bingen E, Bechet S, Derkx V, Werner A, et al. [Nasopharyngeal carriage of children 6 to 60 months during the implementation of the 13-valent pneumococcal conjugate vaccine]. Arch Pediatr [Internet]. 2012 Oct [cited 2014 May 12];19(10):1132–9.
86. da Gloria Carvalho M, Pimenta FC, Jackson D, Roundtree A, Ahmad Y, Millar E V, et al. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. J Clin Microbiol [Internet]. 2010 May [cited 2015 Jul 8];48(5):1611–8.
87. Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kaltoft MS, Reeves PR, et al. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol [Internet]. 2007 Nov [cited 2014 Jun 9];189(21):7841–55.

CHAPTER 4

Streptococcus pneumoniae Serotype Distribution And Dynamics Of Nasopharyngeal Carriage During The First Year Of Life In South African Children

4.1 INTRODUCTION

The pneumococcus, along with a myriad other bacterial species, forms part of the human commensal flora that colonises the nasopharyngeal (NP) mucosal epithelium asymptomatically (1). Children under the age of two years are more frequently colonised by pneumococci than older children and adults (2,3).The nasopharynx therefore serves as a reservoir and source for person-to-person transmission of pneumococci (2-4). Invasive pneumococcal disease (IPD) is an evolutionary dead-end because none of the common forms of pneumococcal disease, promote the transmission of pneumococcus (5). This suggests that pneumococcal virulence determinants are probably evolutionary adaptations that increase its persistence within a host during colonisation (5). Acquisition of pneumococcal carriage can occur as early as the first day of life (6), peaks in children between 2-3 years of age when 50-90% of children may be colonized (3,7). The prevalence subsequently declines to 5 - 10% in older children from developed countries (3,8,9), whereas 25% - 60% of older children from developing countries remain colonised (10-12). Pneumococcal carriage studies are important in understanding the dynamics of pneumococcal serotype-specific colonisation and the effects of vaccination on serotype distribution.

4.1.1 Pneumococcal carriage: A fundamental step in pneumococcal disease

The outcome of NP carriage involves a complex interplay between host-dependent factors such as, age, immune status, underlying medical conditions, pathogen-dependent factors such as virulence characteristics of the colonising strain and co-colonisation and interactions with other microbes. The mechanisms through which

asymptomatic carriage evolves into invasive disease are poorly understood. However, it is believed that the adherence of the pneumococcus to the NP epithelial mucosa via multiple receptor-ligand interactions is the initial step (13,14). Host mucosal immunity including mucus-mediated clearance, serotype-specific antibodies (immunoglobulin G [IgG], secretory IgA, IgM), complement, and opsonophagocytic cells are key to preventing acquisition of pneumococcal carriage (2,3). Escape from immunological control may result in establishment of colonisation, bacterial proliferation and increased colonisation density, which may in turn lead to spread to the surrounding tissue compartments to cause non-invasive mucosal disease (otitis media [OM], sinusitis, pharyngitis), Figure 15. Pneumococci may also be aspirated or spread by direct extension to the lower respiratory tract to cause bronchitis, bronchiolitis and non-bacteraemic pneumonia (3,15). The translocation of the bacterium across mucosal surfaces into the bloodstream may result in septicaemia and/or meningitis (16,17).

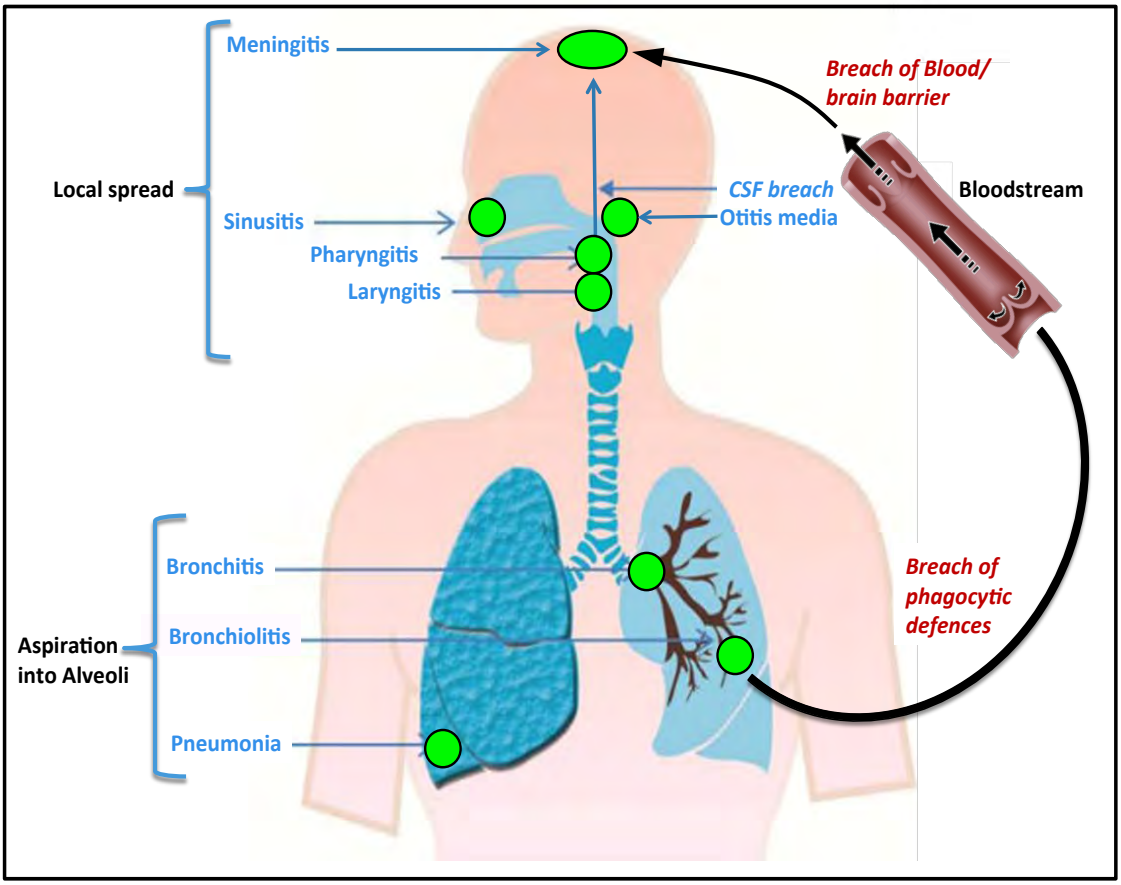

Figure 15: Pneumococcal diseases and pathogenic route. An inoculum of pneumococci via airborne droplets may colonise the nasopharynx and spread local from the nasopharynx into the middle ear, sinus, and larynx leading to otitis media, sinusitis, and laryngitis, respectively. Pneumococci can also be aspirated into alveoli leading to pneumonia or translocate into the bloodstream or meninges and cause septicemia and meningitis respectively. Figure modified from Bogaert *et al* (3)

4.1.2 Serotype-specific acquisition rate and duration of pneumococcal carriage

Pneumococcal disease is associated with recent acquisition of a pneumococcal serotype rather than prolonged carriage duration (15). However, it is during prolonged carriage that exchange of genetic determinants, such as those encoding antibiotic resistance, virulence and serotype, occurs via transformation, enabling the bacterium to overcome selective pressures including antibiotic exposure and immunisation (18). Previous

pneumococcal carriage studies employed monthly NP swab collection strategies with an inherent limitation of underestimating the carriage duration in instances where some serotypes are carried for short durations. Pneumococcal carriage duration varies considerably across epidemiological settings, with carriage duration estimates ranging from 10 days (serotype 3) to 120 days (serotype 6B) (19,20). The duration of pneumococcal carriage is dependent on a number of risk factors including the pneumococcal capsular polysaccharide (CPS), previous exposure to a homologous strain or cross-reactive serotype, age, immune status (19,21), presence of other siblings within the household, prior immunisation and environmental exposures (22–24).

Cross-sectional pneumococcal carriage studies have been key in understanding pneumococcal carriage dynamics pre – and post pneumococcal conjugate vaccine (PCV) implementation (25–28). These studies however fail to provide data on the acquisition and duration of carriage, by virtue of their cross-sectional design. Given the complex interplay between the pneumococcus and other potential carriage determinants (3,29–31), pneumococcal carriage dynamics are best understood from longitudinal data (7,9,11,15,18,19,21,32–37). When the sampling frequency, length of follow-up and clinical data collection are adequate, longitudinal study designs allow the estimation of acquisition, carriage duration and relevant risk factors (9,15,19,34).

4.1.3 Pneumococcal serotype distribution

The pneumococcal sero-epidemiology among NP carriage isolates varies across geographical region, age group, and ethnicity.

Prior to PCV-7 implementation, children from The Gambia were more likely to be colonised by PCV-7 serotypes 6B (14.4%), 19F (10.4%), 14(8.6%), 23F (8.3%), 4 (1.7%) and 9V (2.1%). Serotype 18C was not detected in cohorts from this country. The most

common non-PCV-7 serotypes were serotypes 6A (8.9%), 19A (5.0%), 35B (3.9%), 15B (3.2%), 3 (2.8%), 11 (2.7%) and 16F (2.3%) (34). In Kenya, prior to PCV-7 implementation, children were colonised by serotypes 6B (10.4%), 14 (10.4%), 19F (5.4%), 9 (5.4%), 23F (4.3%) and 4 (1.1%). Serotype 18C was also not detected in this cohort. The most notable non-PCV-7 in this cohort were serotypes 13 (32.6%), 15 (13.0%), 20 (4.3%), 23 (4.3%), 7 (2.2%), 6A (1.1%) and 1 (1.1%) (45). Similarly, South Africa children were more likely (prior to PCV-7 implementation, 2009) to be colonised by PCV-7 serotypes 19F (10%), 6B (13%), 23F (11%), 18C (2%), 14 (15%), 4 (2%), and 9V (3%). Unlike the Kenyan study, serotype 13 was not isolated in any of the South African children that were included in this study (46). Other non-PCV-7 serotypes detected among South African children were serotypes 6A (12%), 19A (7%), 1 (5%), 15 (3.8%), 11 (0.8%), 5 (2%), 3 (0.8%), 7F (0.3%), 20 (0.4%), 16F (0.4%) and 20 (0.4%) (46,47). Such studies provide crucial baseline data for monitoring vaccine efficacy. Although fluctuations in serotype distributions may take place over time even without obvious changes in selective pressures, significant changes in the pneumococcal population structure have been seen post PCV implementation and in relation to the volume of antibiotic prescription (48-50).

Individual serotypes are associated with differing potential to cause invasive disease, antimicrobial drug-resistance and have different global distribution patterns (30,38–41). Of the more than 98 pneumococcal serotypes described to date (42), only a third account for approximately ≥70% invasive pneumococcal disease (IPD), with serotype/-groups 1, 3, 6, 14, 19 and 23 being the most common causes of invasive disease worldwide (43). Prior to the introduction of PCV-7, serotype 14 was noted as the most common cause of IPD in the United States of America, Canada, Europe, Oceania and

Latin America, while serotypes 1, 5 and serogroup 6 were most common causes of IPD in Asia and Africa respectively(3,44). Minor differences in serotype distribution were noted amongst African countries in the pre-PCV-7 era.

4.1.4 Inter-and intra species competition within the NP microbiome: A fitness test of a "sugar-coated commensal"

Studies of the NP microbiome have identified more than 700 bacterial species including *Moraxella spp*, Staphylococcus *aureus, Haemophilus spp., Neisseria spp.,* and *Streptococcus spp* (1). It is believed that inter- and intra species competition within this ecological niche contributes to fluctuations in pneumococcal carriage over time (50,51). Interspecies competitive strategies includes production of bacteriocins (antimicrobial peptides or proteins) which attack other bacteria and are accompanied by an immunity protein to protect against self-destruction (52–55); production of hydrogen peroxide which inhibits the growth of *S. aureus*, *H. influenzae*, *N. meningitidis,* and *Moraxella catarrhalis* (56,57) and secretion of neuraminidase enzyme (NanA) which enzymatically cleaves sialic acid residues from the capsules of *H. influenzae* and *N. meningitidis*, which makes these organisms vulnerable to host defenses (58).

Intra-species competition within pneumococcal populations has been seen following the routine introduction of pneumococcal conjugate vaccine (PCV) (50,59–66). PCVs are efficacious against vaccine-serotype (VT) disease (67–73), discussed in chapter 5. However, non-vaccine-serotype (NVT) disease has increased (60). PCV not only reduces VT disease but is also thought to reduce the NP carriage of VT serotypes (50,74–77). The overall carriage of pneumococci within the nasopharynx has remained largely unchanged due to a parallel increase in the prevalence of NVT serotypes (50). This phenomena occurs either through capsular switching or serotype replacement (loss of

competitive strains [VT] with emergence of previously out-competed strains [NVT]) (28). Our current understanding of serotype-specific inter-dependencies is poorly explored, for example it is still largely unknown whether colonisation by one pneumococcal serotype makes it less likely for that specific individual to acquire a closely related or different pneumococcal serotype? Taken together, inter- and intra species competitive strategies and perturbations of the NP microbiome (either through antibiotics or vaccination) provide selection pressures, resulting in the rapid evolution of this species (2,28,37,50,60). Pneumococcal sero-epidemiology may therefore vary widely within and across different populations. Further, in order to develop more broadly protective vaccines, we need to improve our understanding of the reasons for persistent carriage with VT serotypes as well as the indirect effects of vaccination on carriage with NVT serotypes.

4.2 Aims and Objectives

The overall aim of the work described in this chapter is to understand the dynamics of pneumococcal NP carriage during the first year of life, in South African children enrolled in a PCV-13 vaccinated birth cohort. Specifically, the objectives are to:

1. Determine serotype-specific acquisition rates and the duration of pneumococcal carriage in the nasopharynx of children.
2. Describe the longitudinal patterns of pneumococcal serotype distribution in this cohort.

4.3 MATERIALS AND METHODS

4.3.1 Study setting

This study of 137 infants enrolled between May 29th 2012 and May 31st 2014 was nested within a larger population-based, longitudinal, prospective birth-cohort study, the Drakenstein Child Health Study. It was conducted at two sites in Paarl, a peri-urban area in the Drakenstein sub-district located 60 km from Cape Town, South Africa (78). The study sites were based at TC Newman and Mbekweni clinics, serving a combined population of 200 000 people. The study location is a stable, semi-urban low socioeconomic status community in South Africa, with a high incidence of childhood pneumonia and where conjugate vaccines (*H. Influenzae* type b [HiB] and PCV) are included in the national immunisation program. The 7-valent pneumococcal conjugate vaccine (PCV-7, Prevenar®, Wyeth Pharmaceuticals Inc.) was introduced into the South African Expanded Program on Immunisation (EPI) in April 2009 with no catch-up immunisation (79). PCV-7 was replaced by a 13-valent pneumococcal conjugate vaccine (PCV-13, Prevenar®, Wyeth Pharmaceuticals Inc.) in June 2011 with limited catch-up at 18 months amongst unimmunised children (79). PCV-13 is administered in a 2+1-dosing schedule at 6 weeks, 14 weeks and 9 months of age (79). The residents from both areas are ethnically, linguistically and culturally heterogeneous with TC Newman serving an Afrikaans-speaking, mixed-race (coloured) community while participants from Mbekweni were largely isiXhosa-speaking ethnic black Africans. The details of the study population and study design have been previously described (78). Briefly, questionnaires about maternal health were administered and detailed birth information was obtained at the time of delivery and during each study visit. Trained study nurses performed the interviews in the mother's home language. Data for pneumonia risk

factors and the child's respiratory symptoms were obtained at scheduled visits, results presented in chapter 5. Missed visits were re-booked with a study mobile phone network system or by study community-based fieldworkers.

4.3.2 Participant enrollment

Pregnant women (>18 years) between 20 and 28 weeks' gestation and attending antenatal care at one of the two clinics were enrolled and prospectively followed-up through pregnancy and childbirth (final n=1000 live births, 137 included in the nested study described here). Thereafter, NP swabs were collected every second week during the first year of life as well as during scheduled clinic visits at 6, 12, 18, 24 months of age and during any episode of pneumonia. The collected NP swabs were immediately placed into 1 ml STGG, transported on ice to the laboratory and frozen at -80°C for batch processing as described in chapter 3 Section 3.3-6 (80-82). The figure 16 below, illustrates the overall DCHS study design.

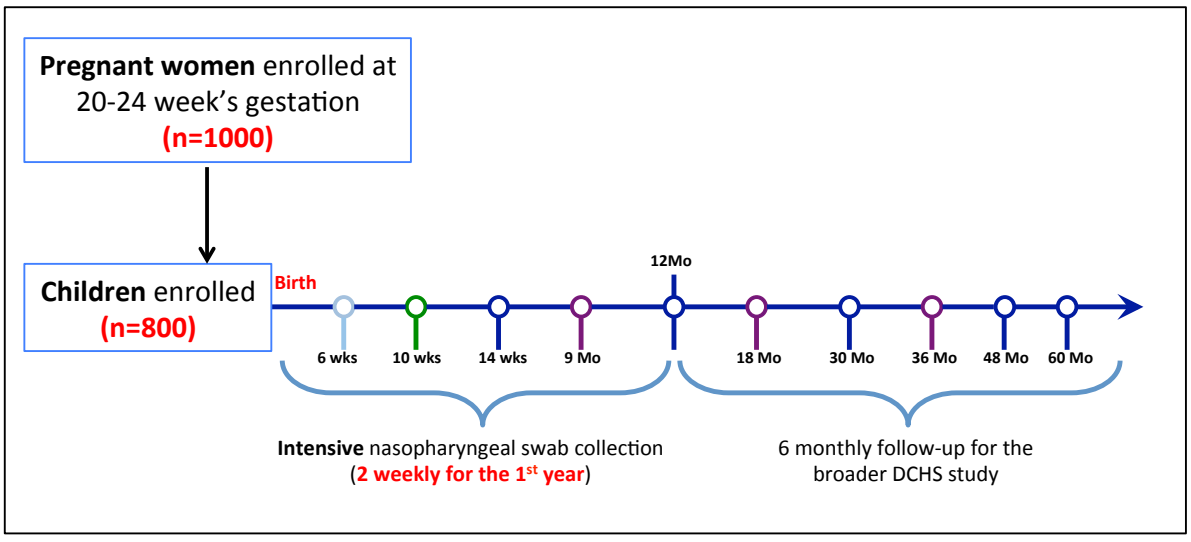

Figure 16: Summary of participant recruitment. A total of 1000 pregnant women were enrolled into the study and followed up for 60 months (5 years of age). A subset of 800 infants will be intensively (2-weekly) followed up for 12 months.

4.3.7 Statistical analysis

Initial exploratory statistics were performed using STATA software (Stata Corporation, College Station, TX) and the rest of the analyses were performed using the openly available statistical environment R, version 3.1.1(83).

Definition of carriage

A new acquisition of pneumococcal carriage was defined when a pneumococcal serotype was cultured from an NP swab for the first time or when a pneumococcal serotype was re-cultured after two consecutive negative NP swabs for that specific, figure 17 below (19,84). A pneumococcal carriage episode, representing ongoing NP colonisation, was therefore defined as the period between acquisition and clearance of the same pneumococcal serotype. Pneumococcal carriage duration was defined as the difference between the clearance date and the date of carriage acquisition. The average duration for each serotype is simply the average of each individual's carriage duration (that is, if a child had two separate carriage episodes with the same serotype, these were regarded as independent events that would be added together and divided by the number of episodes in order to determine the carriage duration).

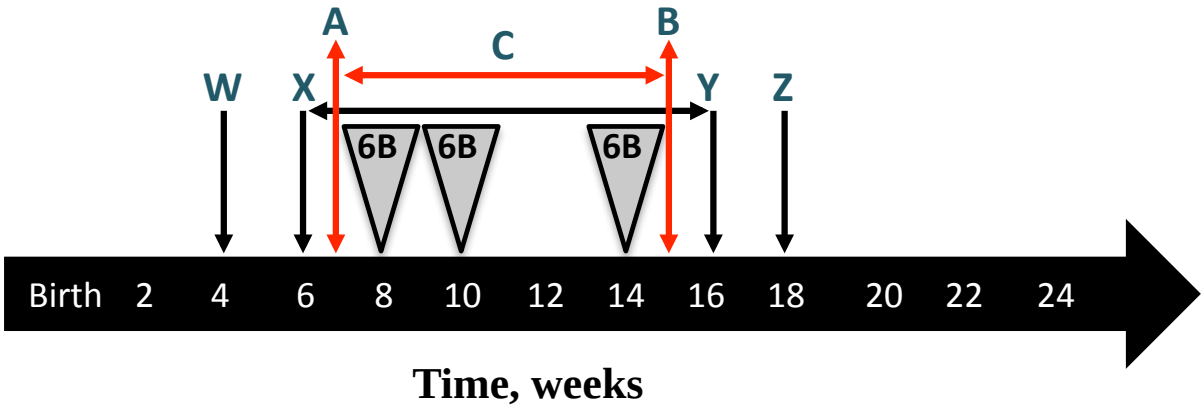

Figure 17: Pneumococcal carriage definition. Black arrows labelled W, X, Y and Z (including week 12) represent sampling points with **negative** NP swabs. Acquisition of serotype 6B pneumococcus, **A**, is defined as the midpoint between the last of two negative (W and X) swabs, in this case, **X** (at week 6) and the first positive NP swab (at week 8). Clearance of carriage, **B**, is defined as the mid-point between the last positive swab for 6B (at week 14) and the first (in this case Y, at week 16) of two consecutive negative swabs (Y and Z). Carriage duration, C, is the difference between the clearance date and date of acquisition (that is B – A), modified from Turner *et al* (19).

Time to pneumococcal acquisition and carriage duration were determined by Kaplan-Meier survival estimates and recurrent colonisation episodes determined by the conditional gap-time model. Survival analysis is the analysis of time to failure. The Cox proportional hazards model, a semi parametric model that allows for no assumptions to be made about the parametric distribution of the survival times, was used. The proportional hazards assumption refers to the fact that the hazard functions are multiplicatively related. Further extensions were made to the Cox model, including investigation of the time to recurring events, when the events are not independent. Children are colonised with pneumococci repeatedly through their first few years of life. One such model to investigate the time to recurring events is the model developed by Prentice, Williams & Peterson (1981), commonly known as the conditional gap time model. The model investigates time to failure from the subject's previous time of failure. Here the survival function measures the probability of not experiencing the second

event before time$t_2 - t_1$ on condition that the individual experienced the first failure at timet_1, wheret_1 andt_2 are the failure times of the first and second events respectively

4.3.8 Ethical considerations

The study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town (HREC ref: 401/2009 and 740/2013) and the Western Cape Provincial Child Health Research Committee. Mothers provided written informed consent at enrolment and annually thereafter.

4.4 RESULTS

4.4.1 Demographic characteristics

One hundred and thirty seven infants (median age at censoring, 365 days; interquartile range, IQR 363 – 366) were enrolled in the present study. A total of 3331 NP swabs were collected and pneumococci were isolated from 1809 of the 3331(54%) swabs collected, figure 18. Twenty eight (28) isolates were not viable on subculture hence not included in the present analysis. Using sequetyping, we were unable to amplify the *wzh* gene in 172/1809(9.5%) isolates as confirmed by Quellung testing. A single amplicon of ~ 1,061 bp was detected in 1637/1809 (90%) isolates of which 1536/1637 (94%) were correctly assigned a serotype. An additional 101 isolates with ambiguity were assigned a serotype by Quellung testing. The overall number of isolates correctly serotyped was therefore 1637/1809 (90%).

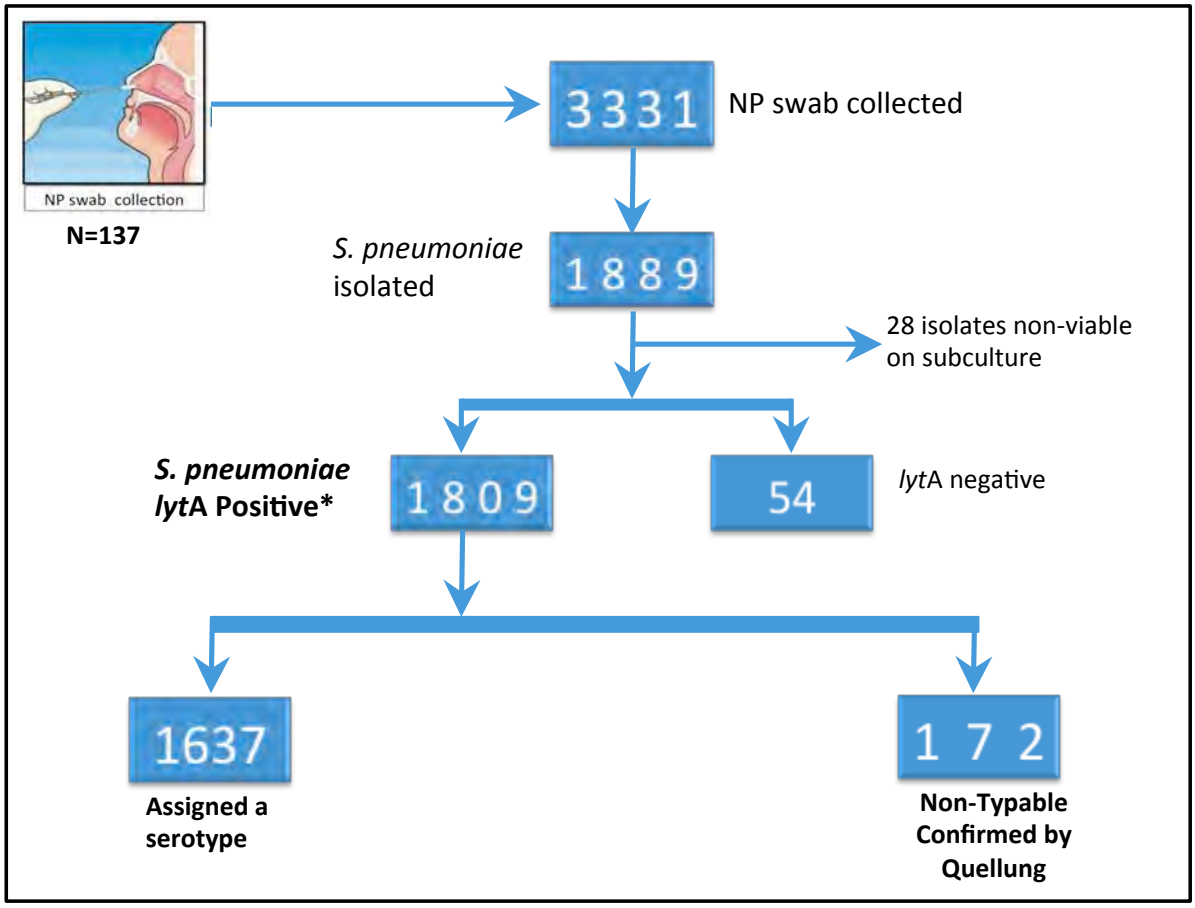

Figure 18: Summary of pneumococcal isolates recovered from 3331 nasopharyngeal (NP) swabs collected from 137 longitudinally followed up children.

4.4.2 Pneumococcal acquisition dynamics

Figure 19 shows the survival curve estimates, by age, for the first acquisition of a pneumococcal strain. All but seven children were colonised at least once by 260 days of life. Three children were never colonised during their first year of life. By 120 days, 75% of the children had been colonised by a pneumococcus at least once. Given our observation above that a child can have recurrent pneumococcal carriage acquisition episodes, we applied the conditional-gap time model to account for the multiple failure events, figure 20. Children often took longer to acquire their first carriage episode but once they had, subsequent colonisation episodes occurred much faster, figure 20 and

table 13. The difference between time to first pneumococcal acquisition and the time to recurrent episodes was statistically significant, figure 20 and table 13. The median time to first pneumococcal acquisition was 63 days (95% CI 55 – 90 days), with no difference observed between VT and NVT isolates (p = 0.698).

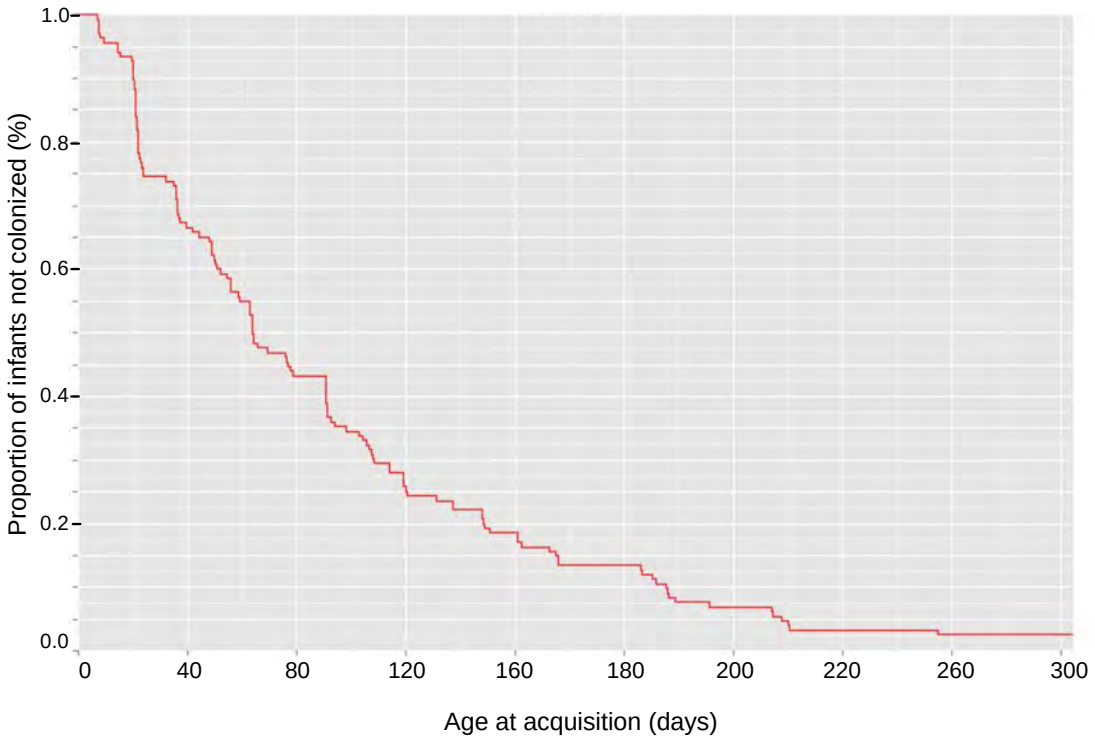

Figure 19: Kaplan Meier survival estimate showing the time to first pneumococcal carriage episode in a cohort of South African children during their first year of life, n=137.

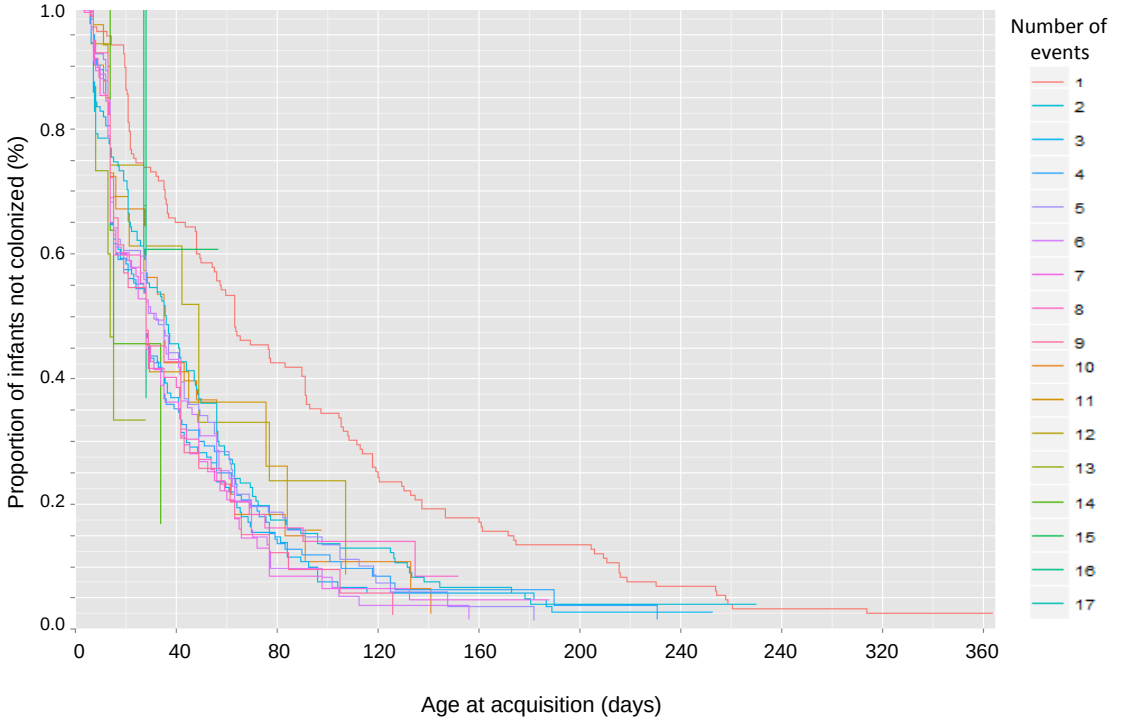

Figure 20: Conditional Gap model for recurrent pneumococcal acquisition events during the first year of life in children participating in a birth cohort, n=137.

Table 13: Conditional Gap model for median time to recurrent pneumococcal acquisition episodes.

Recurrent Acquisitions*	Number of children	Number of episodes	Median (days)	95% CI
1st	137	134	63	55 - 90
2nd	134	128	36	28 - 47
3rd	127	122	28	21 - 35
4th	122	113	28	19 - 35
5th	111	102	31	28 - 42
6th	98	89	32	22 - 42
7th	84	73	28	21.5 - 35
8th	68	56	28	15 - 42
9th	51	42	28	17 - 42
10th	40	34	35	28 - 56
11th	30	19	28	21 - NA
12th	15	11	49	27 - NA
13th	7	5	14	8 - NA
14th	5	4	15	14 - NA
15th	2	1	NA	27 - NA
16th	1	1	28	28 - NA
17th	1	0	NA	NA - NA

*1st recurrent acquisition implies 2nd carriage acquisition episode.

4.4.3 Pneumococcal carriage prevalence

None of the children enrolled in the present study were colonised at birth, figure 21. The pneumococcal point prevalence at 2 weeks of age was 2%, reaching a maximum prevalence of 60% at 24 weeks of age, thereafter plateauing around 55% for the remainder of the first year of life. When examined for trend, the pneumococcal point prevalence increased linearly with time (X^2 test for trend p = 0.0001). At each sampling point, non-PCV-13 serotypes accounted for more than 80% of the pneumococci that were isolated.

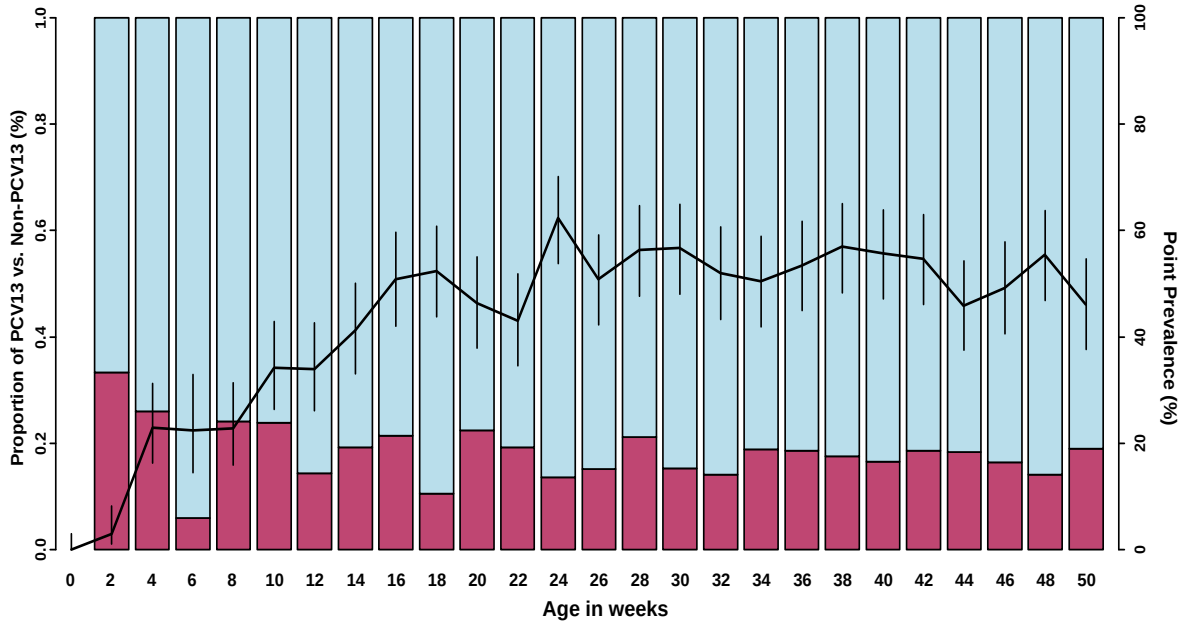

Figure 21: Pneumococcal point prevalence showing the number of children colonised at each sampling point and the proportion of vaccine type or non-vaccine type pneumococci detected. The point prevalence of carriage increased linearly with time after birth. Red bars = PCV-13 serotype, Blue bars = Non-PCV-13 serotypes.

4.4.4 Pneumococcal serotype distribution

Of the 48 different pneumococcal serotypes detected in this cohort, the most frequently encountered VT serotypes included 19F, 9V, 19A and 6A while NVT serotypes included

15B/15C, 21, 10A, 16F, 35B, 9N and 15A, figure 22. Of the pneumococcal serotypes detected in this study, 1268/1809 (70%) were not included in the current PCV-13.

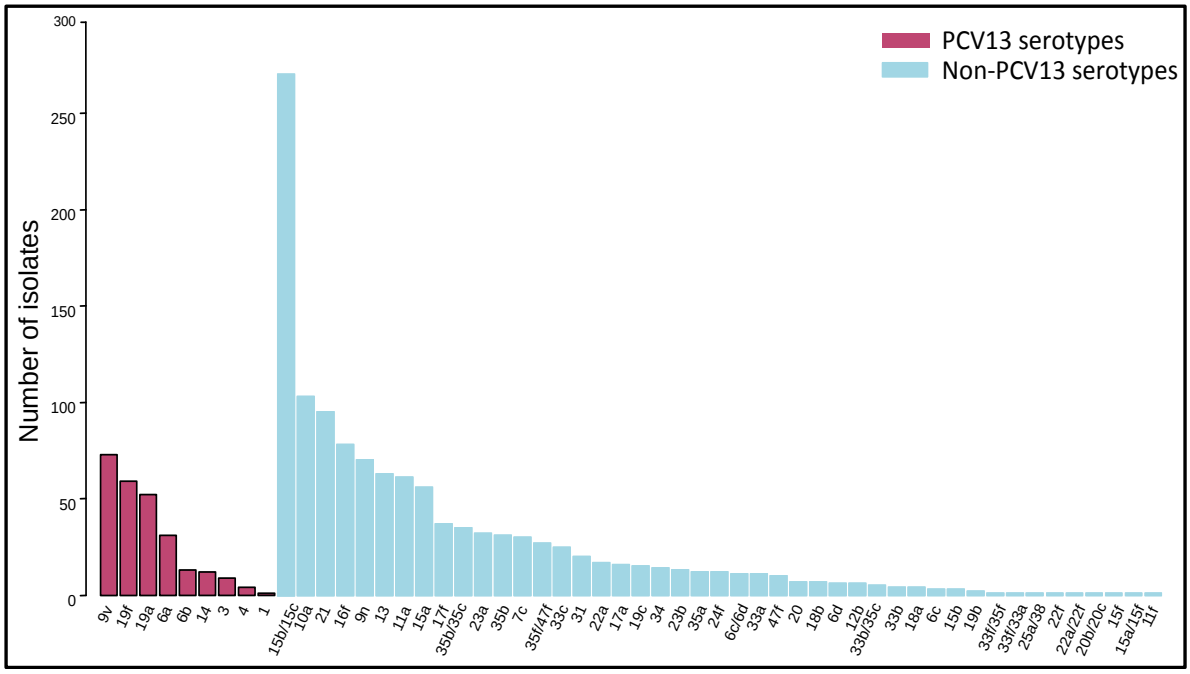

Figure 22: Pneumococcal serotype distribution. Red bars = PCV-13 serotype, Blue bars = Non-PCV-13 serotypes. Serotype 33F and 22F are included in the PCV-15 vaccine formulation that is being evaluated in clinical trials.

Figure 23 – 25 shows the longitudinal carriage patterns for all the 48 different pneumococcal serotypes encountered in the present study. There was a slight drop in the prevalence of VT (serotypes 9V, 19F, 19A, 6A, 6B, 14, 3, 4 and 1) at weeks, 6 and 14 which coincides with the first and second scheduled dose of PCV-13 immunisation in our cohort, figure 23. Thereafter, the overall longitudinal distribution of VT seems to plateau after week 14. Similar trends where seen for the 10 most predominant NVT (Serotypes 15B/15C, 10A, 21, 16F, 9N, 13, 15A, 35B/35C, 17F and 11A/11D) with an initial sharp increase after 6 weeks, and plateau after 14 weeks, figure 24. These patterns were however not consistent in the other less frequently encountered NVT, figure 25.

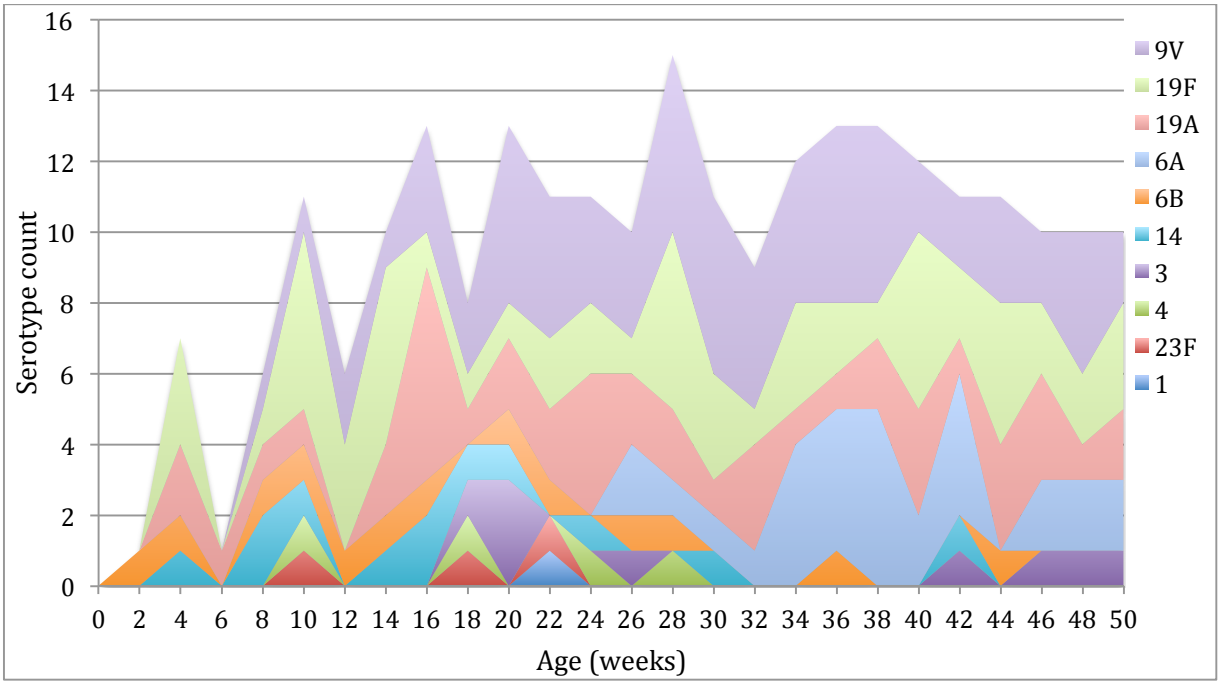

Figure 23: Longitudinal carriage patterns of PCV-13 vaccine (VT) serotypes during the first year of life, n=137. The number of pneumococci detected at each sampling point is shown by the thickness of each mountain graph, that is, at 16, 28, 36 and 48weeks, 3, 5 and 4 children were colonised by serotype 9V respectively.

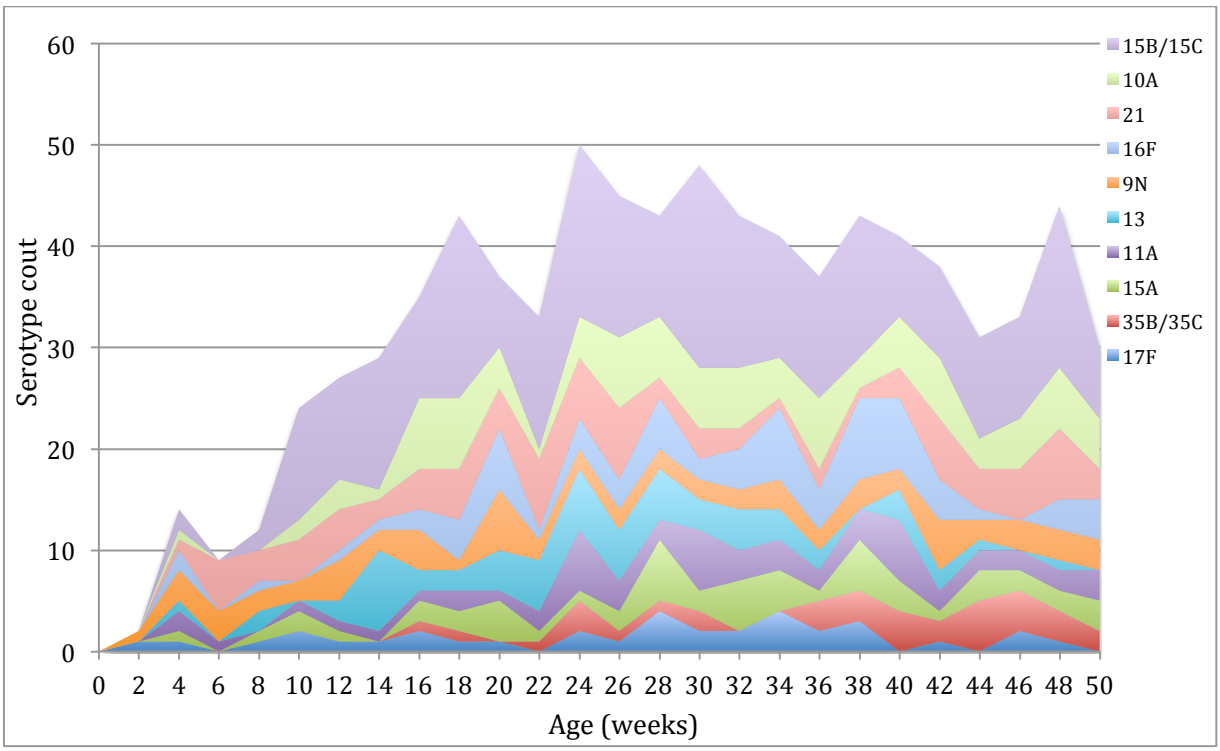

Figure 24: Longitudinal carriage patterns of 10 most predominant non-PCV-13 vaccine (NVT) serotypes during the first year of life, n=137. The number of pneumococci detected at each

sampling point is shown by the thickness of each mountain graph, that is, at 16, 28, 36 and 48 weeks, approximately 10, 12, 10 and 5 children were colonised by serotype 15B/15C respectively.

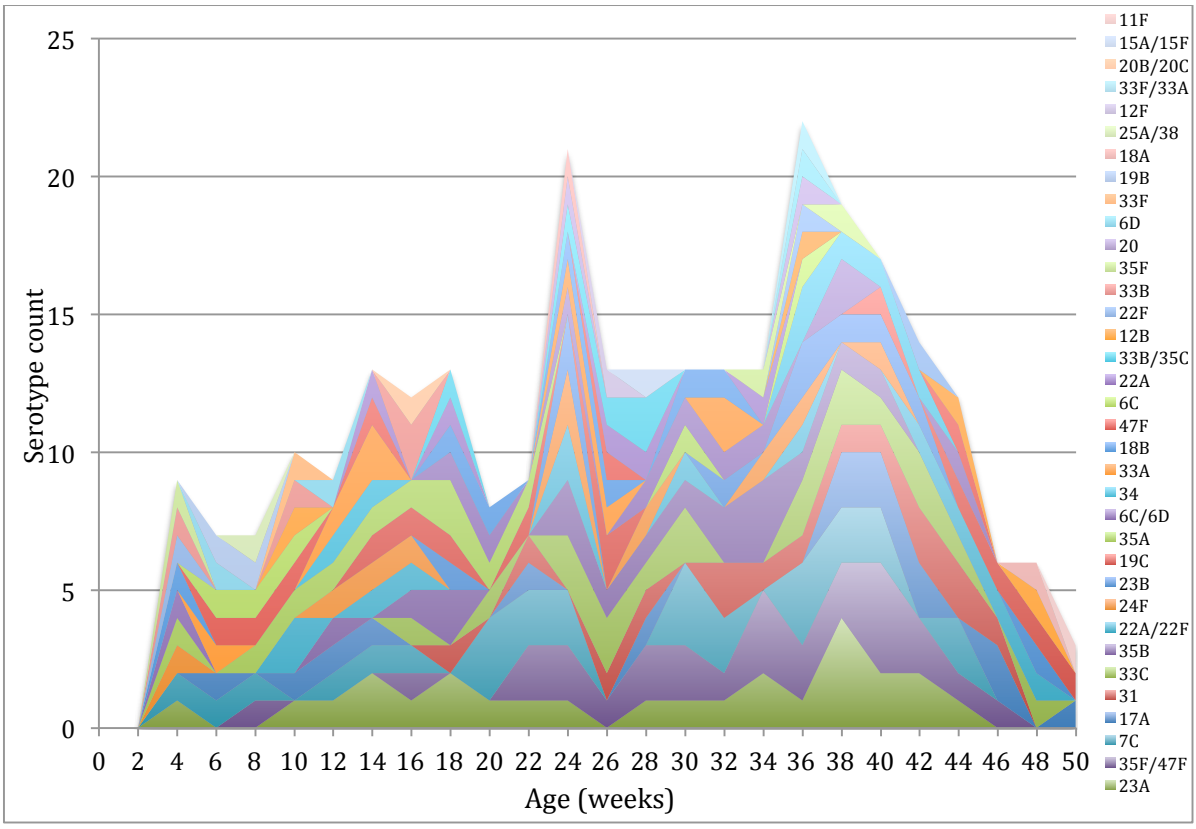

Figure 25: Longitudinal carriage patterns of all the remaining non-PCV-13 vaccine (NVT) serotypes during the first year of life, n=137. The number of pneumococci detected at each sampling point is shown by the thickness of each mountain graph. Most of these serotypes were uncommon and more often detected in one child per scheduled visit.

4.4.5 Pneumococcal serotype-specific acquisition and carriage duration

Serotype-specific acquisition rates ranged from 0.0075 to 0.8827-episodes/child year (e/cy), whilst the carriage duration for the 48 different serotypes detected was between 0.04 and 29.42 days. Collectively, the rate of VT, NVT and non-typeable pneumococcal were 1.07 e/cy (95% CI 0.91 -1.26), 5.10 e/cy (95% CI 4.74 - 5.49) and 0.815 e/cy (95% CI 0.68 - 0.98) respectively, table 15.

Table 14: Rate of pneumococcal acquisition and duration of carriage among infants, by serotype, in decreasing order of number of acquisitions.

Serotype	Number of new Acquisitions	Rate of acquisition[b]	95% CI	Carriage Duration[a]	95% CI
All VT	**143**	**1.07**	**0.91 - 1.26**	**28.31**	**20.35 - 36.28**
9V*	33	0.24	0.17 - 0.34	7.82	3.03 - 12.62
19F*	29	0.21	0.15 - 0.31	6.53	3.10 - 9.96
19A*	30	0.22	0.15 - 0.32	5.39	2.29 - 8.48
6A*	18	0.13	0.08 - 0.21	3.51	1.20 - 5.81
14*	7	0.05	0.02 - 0.11	1.47	0.14 - 2.80
6B*	4	0.03	0.01 - 0.08	1.60	0.00 - 4.16
3*	5	0.04	0.01 - 0.09	0.92	0.00 - 1.89
4*	4	0.03	0.01 - 0.00	0.41	0.00 - 0.89
1*	1	0.01	0.00 - 0.05	0.10	0.00 - 0.30
All Non-VT	**682**	**5.10**	**4.74 - 5.49**	**129.04**	**115.48 - 142.61**
15B/15C	118	0.88	0.73 - 1.05	29.44	21.97 - 36.89
10A	42	0.31	0.23 - 0.42	10.81	6.09 - 15.52
21	45	0.33	0.25 - 0.45	9.75	5.55 - 13.96
16F	36	0.26	0.19 - 0.37	8.21	4.32 - 12.10
9N	33	0.24	0.17 - 0.34	7.20	3.61 - 10.79
13	27	0.20	0.13 - 0.29	6.58	2.93 - 10.24
11A	34	0.25	0.18 - 0.35	5.99	3.05 - 8.93
15A	31	0.23	0.16 - 0.33	5.94	2.51 - 9.36
17F	24	0.18	0.12 - 0.26	4.05	1.49 - 6.61
35B/35C	25	0.18	0.12 - 0.27	3.51	1.90 - 5.11
23A	14	0.10	0.06 - 0.17	3.15	0.83 - 5.47
35B	25	0.18	0.12 - 0.27	2.94	1.77 - 4.11
7C	19	0.14	0.09 - 0.22	2.88	1.09 - 4.66
35F/47F	9	0.06	0.03 - 0.12	2.78	0.04 - 5.52
33C	14	0.10	0.06 - 0.17	2.36	0.40 - 4.32
31	10	0.07	0.04 - 0.13	2.29	0.23 - 4.36
22A	10	0.07	0.04 - 0.13	1.86	0.20 - 3.52
17A	12	0.09	0.05 - 0.15	1.60	0.54 - 2.67
24F	12	0.09	0.05 - 0.15	1.38	0.57 - 2.20
6C/6D	7	0.05	0.02 - 0.11	1.35	0.00 - 3.01
19C	7	0.05	0.02 - 0.11	1.33	0.04 - 2.63
35A	3	0.02	0.01 - 0.07	1.31	0.00 - 3.54
34	9	0.06	0.03 - 0.12	1.30	0.27 - 2.32
23B	6	0.04	0.02 - 0.10	1.29	0.00 - 3.14
33A	10	0.07	0.04 - 0.13	1.08	0.24 - 1.92
47F	7	0.05	0.02 - 0.11	0.96	0.21 - 1.72
20	6	0.04	0.02 - 0.10	0.67	0.00 - 1.38
18B	5	0.03	0.01 - 0.09	0.66	0.00 - 1.38
12B	6	0.04	0.02 - 0.10	0.55	0.10 - 0.99
33B/35C	4	0.03	0.01 - 0.08	0.45	0.01 - 0.90
6D	4	0.03	0.01 - 0.08	0.44	0.00 - 0.98
33B	4	0.03	0.01 - 0.08	0.40	0.01 - 0.79
15B	3	0.02	0.01 - 0.07	0.31	0.00 - 0.65
18A	2	0.02	0.00 - 0.06	0.30	0.00 - 0.74
6C	2	0.02	0.00 - 0.06	0.25	0.00 - 0.75
19B	2	0.02	0.00 - 0.06	0.24	0.00 - 0.57
15A/15F	1	0.01	0.00 - 0.05	0.10	0.00 - 0.32
11F	1	0.01	0.00 - 0.05	0.10	0.00 - 0.31
25A/38	1	0.01	0.00 - 0.05	0.10	0.00 - 0.31
22A/22F	1	0.01	0.00 - 0.05	0.10	0.00 - 0.30
33F/33A	1	0.01	0.00 - 0.05	0.10	0.00 - 0.30
33F/35F	1	0.01	0.00 - 0.05	0.10	0.00 - 0.30
20B/20C	1	0.01	0.00 - 0.05	0.09	0.00 - 0.29
22F	1	0.01	0.00 - 0.05	0.09	0.00 - 0.29
15F	1	0.01	0.00 - 0.05	0.04	0.00 - 0.13
Nontypeable	**109**	**0.815**	**0.67 - 0.98**	**16.66**	**13.59 - 19.73**

[a]Duration of carriage in days, [b]Acquisition rates given as episodes per child year.
*Serotypes included in PCV-13 vaccine formulation.

Similarly, the carriage durations were 28.31 days (95% CI 20.35 - 36.28), 129.04 days (95% CI 115.48 - 142.61) and 16.66 days (95% CI 13.59 - 19.73) for VT, NVT and non-

typeable pneumococcal respectively. We applied a one-way analysis of variance to investigate the difference between carriage duration for VT, NVT and non-typeable pneumococci. The analysis further showed a significant difference in at least one of the three pneumococcal groups ($p < 0.001$). A Bonferroni post hoc test yields a significant difference in carriage duration for VT and NVT pneumococci, $p< 0.0001$, and also a significant difference between NVT and non-typeable pneumococci, $p < 0.0001$. There was no significant difference ($p = 1.000$) in the carriage duration between VT and non-typeable pneumococci.

4.5 DISCUSSION

This is the largest African study that longitudinally describes the dynamics of pneumococcal carriage over the first year of life, both in terms of the number of intensively sampled participants and the collection of detailed clinical phenotypes. Despite the high immunisation coverage (with 94%, 94%, 92% and 89% coverage at 6, 10 and 14 weeks and 9 months visits respectively), pneumococcal carriage prevalence (>55%) and carriage duration of 48 serotypes in this population remained high. In terms of carried pneumococci, children were less likely to be colonised by VT pneumococci compared to NVT (point prevalence >80% throughout the sampling period). This is consistent with other studies that have shown similar proportions of NVT post PCV-13 implementation which range from 73% to 97% (59,66,87–92). Once they acquired their first pneumococcal carriage episode, children experienced subsequent carriage episodes much earlier. We further show that, as the children grew older, the pneumococcal point prevalence also increased with age (X^2 for trend $p = 0.0001$) over the first year of life.

The median age to first pneumococcal acquisition (63 days) is twice as high compared reports from other similar LMIC settings. The median age to first acquisition in the Gambia was reported as 24 days (34) which is lower than 38.5 days in Kenya. Both of these were baseline studies conducted prior to the introduction of PCV-7 into their national immunisation schedules. None of the children included in our cohort were colonised at birth and acquired a pneumococci less rapidly than in other settings. This may be a function of herd protection due to the high immunisation coverage, or it may simply suggest that children in our settings are protected against acquisition either by passive maternal antibody or by behavioral factors limiting their exposure to infection.

Children are often colonised by several different pneumococcal serotypes over the first years of life, and the less immunogenic serotypes tend to be carried within the nasopharynx for prolonged periods of time (e.g., 6, 14, 19, and 23) than the more immunogenic strains (e.g., 3, 12, and 33) (93). The longer duration of carriage observed in our settings for serotypes 9V, 19F, 19A, 6A, 6B and 14 are consistent with these findings (carriage durations ranging from 2 to 8 days). Serotypes 3, 4 and 1 were carried for shorter periods of times (carriage duration 0.99, 0.42, 0.10 days respectively). In contrast however, NVT were generally carried for longer periods of time than VT. In particular, serotypes 15B and 15C were carried for approximately 29 days on average and were more frequently acquired, 0.88 e/cy. Other notable NVT that were carried for prolonged periods of time include serotypes 10A, 21, 16F, 9N, 13, 11A, 15A and 17F.

Serotypes such as 15B, 15C, 10A, 21, 16F, 9N, 13, 11A, 15A, 17F, 31 and 22A are currently among the most prevalent replacement NVTs globally post PCV-13 implementation (34,64,66,89). NP colonisation increases the risk of disease (3). The observation of the near elimination of PCV-7 serotypes and increase in non-PCV-7

serotypes has led to the formulation of higher valency PCV vaccines, the most recent being PCV-13 and PCV-15 (currently under development). In light of the disproportionate increase in NVT observed in our study and others, the potential of these NVT in causing both mucosal and invasive pneumococcal disease constantly overshadows the success of current PCV. It is unknown whether other serotypes will continue to cause replacement disease until all 98 serotypes have had their moment of fame.

The introduction of PCV-7 into nationwide routine immunisation programs has had a significant impact in reducing the burden of pneumococcal disease in many settings but these strategies have also been accompanied by significant changes in the pneumococcal population structure. In many developing countries, serotypes 1 and 5 soon became important cause of IPD post PCV-7 licensure; in South Africa serotype 1 caused approximately 13.5% of IPD cases in children and were subsequently included in PCV9 formulations (44,62,65,93). Similar observations were made for serotype 7F and serotype 3 which became incorporated into PCV-10 and PCV-11 respectively (86,94,95). The public health benefits of PCV were further offset by the demonstration that serotypes 6B and 19F included in PCV formulations offered no cross-protection against closely related serotype 6A and 19A IPD cases and hence their inclusion in PCV-13, licensed in the year 2010. Therefore, the widespread use of a new PCV has inevitably led to newer replacement strains, rendering the serotype-specific formulation of PCV a moving target (96,97). Most recently, preliminary data from developed countries suggests that IPD caused by serotypes 22F among children < 5 years of age and adults ≥ 65 years of age were 12.4% and 12.5%, respectively, while serotype 33F was responsible for 7.9% and 2.7% of IPD cases in children < 5 years of age and adults ≥ 65 years of age, respectively (60,74,75,98). In order to address the potential risk associated

with these emerging serotypes, the safety, tolerability and immunogenicity of a 15-valent PCV (PCV-15; Merck & Co. Inc) which contains all PCV-13 serotypes and two additional serotypes, 22F and 33F, has been evaluated in healthy toddlers (99).

Given the demonstrable differences in the serotype distribution between developed and developing countries, the public health benefit PCV-15 in developing countries is uncertain. Long-term surveillance in developing countries is critical in determining the impact of newer PCV formulations on IPD, carriage, and invasive disease. Other challenges include optimization of current PCV formulations in other vulnerable risk groups and ensuring that PCV immunisation is globally accessible.

The clinical and public health implication of co-colonisation by multiple pneumococcal serotypes is not well established. Carriage of more than one pneumococcal serotype has been shown to be common especially in settings where the pneumococcal carriage rate and the burden of disease are high (36,94–97). Our study was however not able to detect multiple colonising pneumococcal because of the technique employed. As described in detail in the preceding chapter, finding a pneumococcal capsular typing technique that is both sensitive and applicable for high throughput typing remains a challenge. Given the anticipated perturbations of the NP microbiome induced by PCV, it is very important that alternative typing techniques that are both sensitive and specific are employed in accurate identification of co-colonising pneumococcal serotypes especially NVT as well as differentiating non-typeable pneumococci from other viridans group streptococci. Both NVT and non-typeable pneumococci are increasing becoming important in disease as well as serve as a pool for virulence genes including antibiotic resistance genes (50,60,98–100).

Simoès *et al.* (111) has recently shown that *lytA*-based identification methods can misidentify *S. pneumoniae*. The 54 isolates that were phenotypically presumed to be

pneumococcus (colony morphology, optochin sensitive and alpha hemolysis) but *lytA* negative need to be investigated further by to determine whether they are in fact pneumococci, for example by genome sequencing and phenotypic testing. Another limitation of the data presented here is the lack of pneumococcal carriage data from the mothers as well as other siblings living in the same household. This information is key in assessing transmission probabilities given the high HIV prevalence in our settings. Our understanding of the pneumococcal colonisation patterns in children born to HIV-infected mothers is incomplete. HIV infected mothers have been shown to have a higher prevalence and rate of overall and VT pneumococcal colonisation, potentially serving as a VT pneumococcal transmission source (64,84,101–103). HIV exposed infants may as well be susceptible to pneumococcal infection due to the reduced or impaired maternal antibodies and therefore acquire VT pneumococci from older siblings.

4.6 REFERENCES

1. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. J Clin Microbiol [Internet]. 2005 Nov [cited 2015 Aug 19];43(11):5721–32.
2. Simell B, Auranen K, Käyhty H, Goldblatt D, Dagan R, O'Brien KL. The fundamental link between pneumococcal carriage and disease. Expert Rev Vaccines [Internet]. 2012 Jul [cited 2015 Feb 20];11(7):841–55.
3. Bogaert D, De Groot R, Hermans PWM. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis [Internet]. 2004 Mar [cited 2014 May 1];4(3):144–54.
4. Kllyhty H, Auranen K, Dagan R, Goldblatt D, O'Brien KL. Case for Carriage. Vaccine [Internet]. 2013 Dec [cited 2015 Feb 20];32(1):180–203.
5. Weiser JN. The pneumococcus: why a commensal misbehaves. J Mol Med (Berl) [Internet]. 2010 Feb [cited 2015 Sep 7];88(2):97–102.
6. Tigoi CC, Gatakaa H, Karani A, Mugo D, Kungu S, Wanjiru E, et al. Rates of acquisition of pneumococcal colonization and transmission probabilities, by serotype, among newborn infants in Kilifi District, Kenya. Clin Infect Dis [Internet]. 2012 Jul [cited 2015 Mar 4];55(2):180–8.
7. Watson K, Carville K, Bowman J, Jacoby P, Riley TV, Leach AJ, et al. Upper respiratory tract bacterial carriage in Aboriginal and non-Aboriginal children in a semi-arid area of Western Australia. Pediatr Infect Dis J [Internet]. 2006 Sep [cited 2015 Mar 4];25(9):782–90.
8. Jounio U, Juvonen R, Bloigu A, Silvennoinen-Kassinen S, Kaijalainen T, Kauma H, et al. Pneumococcal carriage is more common in asthmatic than in non-asthmatic young men. Clin Respir J [Internet]. 2010 Oct [cited 2015 Mar 4];4(4):222–9.
9. Hussain M, Melegaro A, Pebody RG, George R, Edmunds WJ, Talukdar R, et al. A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. Epidemiol Infect [Internet]. 2005 Oct [cited 2015 Mar 4];133(5):891–8.
10. Adetifa IMO, Antonio M, Okoromah CAN, Ebruke C, Inem V, Nsekpong D, et al. Pre-vaccination nasopharyngeal pneumococcal carriage in a Nigerian population: epidemiology and population biology. PLoS One [Internet]. 2012 Jan [cited 2015 Feb 7];7(1):e30548.
11. Kwambana BA, Barer MR, Bottomley C, Adegbola RA, Antonio M. Early acquisition and high nasopharyngeal co-colonisation by *Streptococcus pneumoniae* and three respiratory pathogens amongst Gambian new-borns and infants. BMC Infect Dis [Internet]. 2011 Jan [cited 2015 Jan 29];11:175.
12. Roca A, Hill PC, Townend J, Egere U, Antonio M, Bojang A, et al. Effects of community-wide vaccination with PCV-7 on pneumococcal nasopharyngeal carriage in the Gambia: a cluster-randomized trial. PLoS Med [Internet]. 2011 Oct [cited 2014 May 11];8(10):e1001107.

13. Reddy MS, Bernstein JM, Murphy TF, Faden HS. Binding between outer membrane proteins of nontypeable *Haemophilus influenzae* and human nasopharyngeal mucin. Infect Immun [Internet]. 1996 Apr [cited 2015 Sep 7];64(4):1477–9.

14. Beisswenger C, Lysenko ES, Weiser JN. Early bacterial colonization induces toll-like receptor-dependent transforming growth factor beta signaling in the epithelium. Infect Immun [Internet]. 2009 May [cited 2015 Mar 4];77(5):2212–20.

15. Gray BM, Turner ME, Dillon HC. Epidemiologic studies of *Streptococcus pneumoniae* in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. Am J Epidemiol [Internet]. 1982 Oct [cited 2015 Feb 20];116(4):692–703.

16. Filippidis A, Fountas KN. Nasal lymphatics as a novel invasion and dissemination route of bacterial meningitis. Med Hypotheses [Internet]. 2009 Jun [cited 2015 Feb 20];72(6):694–7.

17. Marra A, Brigham D. *Streptococcus pneumoniae* causes experimental meningitis following intranasal and otitis media infections via a nonhematogenous route. Infect Immun [Internet]. 2001 Dec [cited 2015 Feb 20];69(12):7318–25.

18. Sleeman KL, Griffiths D, Shackley F, Diggle L, Gupta S, Maiden MC, et al. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. J Infect Dis [Internet]. 2006 Sep 1 [cited 2015 Aug 27];194(5):682–8.

19. Turner P, Turner C, Jankhot A, Helen N, Lee SJ, Day NP, et al. A longitudinal study of *Streptococcus pneumoniae* carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. PLoS One [Internet]. 2012 Jan [cited 2015 Mar 4];7(5):e38271.

20. Abdullahi O, Karani A, Tigoi CC, Mugo D, Kungu S, Wanjiru E, et al. Rates of acquisition and clearance of pneumococcal serotypes in the nasopharynges of children in Kilifi District, Kenya. J Infect Dis [Internet]. 2012 Oct 1 [cited 2015 Mar 4];206(7):1020–9.

21. Hill PC, Townend J, Antonio M, Akisanya B, Ebruke C, Lahai G, et al. Transmission of *Streptococcus pneumoniae* in rural Gambian villages: a longitudinal study. Clin Infect Dis [Internet]. 2010 Jun 1 [cited 2015 Mar 4];50(11):1468–76.

22. Huang SS, Finkelstein JA, Lipsitch M. Modeling community- and individual-level effects of child-care center attendance on pneumococcal carriage. Clin Infect Dis [Internet]. 2005 May 1 [cited 2015 Mar 4];40(9):1215–22.

23. Levine OS, Farley M, Harrison LH, Lefkowitz L, McGeer A, Schwartz B. Risk factors for invasive pneumococcal disease in children: a population-based case-control study in North America. Pediatrics [Internet]. 1999 Mar [cited 2015 Mar 4];103(3):E28.

24. Takala AK, Jero J, Kela E, Rönnberg PR, Koskenniemi E, Eskola J. Risk factors for primary invasive pneumococcal disease among children in Finland. JAMA [Internet]. 1995 Mar 15 [cited 2015 Mar 4];273(11):859–64.

25. Cohen R, Levy C, Bingen E, Bechet S, Derkx V, Werner A, et al. [Nasopharyngeal

carriage of children 6 to 60 months during the implementation of the 13-valent pneumococcal conjugate vaccine]. Arch Pediatr [Internet]. 2012 Oct [cited 2014 May 12];19(10):1132–9.

26. Vu HTT, Yoshida LM, Suzuki M, Nguyen HAT, Nguyen CDL, Nguyen ATT, et al. Association between nasopharyngeal load of *Streptococcus pneumoniae,* viral coinfection, and radiologically confirmed pneumonia in Vietnamese children. Pediatr Infect Dis J [Internet]. 2011 Jan [cited 2015 Jun 19];30(1):11–8.

27. Berkley JA, Munywoki P, Ngama M, Kazungu S, Abwao J, Bett A, et al. Viral etiology of severe pneumonia among Kenyan infants and children. JAMA [Internet]. 2010 May 26 [cited 2014 Nov 18];303(20):2051–7.

28. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog [Internet]. 2007 Nov [cited 2015 Jan 23];3(11):e168.

29. Rudan I, O'Brien KL, Nair H, Liu L, Theodoratou E, Qazi S, et al. Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. J Glob Health [Internet]. 2013 Jun [cited 2014 May 1];3(1):010401.

30. Sjöström K, Spindler C, Ortqvist A, Kalin M, Sandgren A, Kühlmann-Berenzon S, et al. Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. Clin Infect Dis [Internet]. 2006 Feb 15 [cited 2015 Sep 24];42(4):451–9.

31. Walaza S, Tempia S, Dawood H, Variava E, Moyes J, Cohen AL, et al. Influenza virus infection is associated with increased risk of death amongst patients hospitalized with confirmed pulmonary tuberculosis in South Africa, 2010¿2011. BMC Infect Dis [Internet]. 2015 Jan 27 [cited 2015 Feb 28];15(1):26.

32. Granat SM, Mia Z, Ollgren J, Herva E, Das M, Piirainen L, et al. Longitudinal study on pneumococcal carriage during the first year of life in Bangladesh. Pediatr Infect Dis J [Internet]. 2007 Apr [cited 2015 Sep 3];26(4):319–24.

33. Hill PC, Akisanya A, Sankareh K, Cheung YB, Saaka M, Lahai G, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers. Clin Infect Dis. 2006 Sep;43(6):673–9.

34. Hill PC, Cheung YB, Akisanya A, Sankareh K, Lahai G, Greenwood BM, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian infants: a longitudinal study. Clin Infect Dis [Internet]. 2008 Mar 15 [cited 2015 Sep 11];46(6):807–14.

35. Madhi SA, Izu A, Nunes MC, Violari A, Cotton MF, Jean-Philippe P, et al. Longitudinal study on *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* nasopharyngeal colonization in HIV-infected and -uninfected infants vaccinated with pneumococcal conjugate vaccine. Vaccine [Internet]. 2015 Apr 21 [cited 2015 Apr 27];

36. Turner P, Hinds J, Turner C, Jankhot A, Gould K, Bentley SD, et al. Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. J Clin Microbiol [Internet]. 2011 May [cited 2014 May 11];49(5):1784–9.

37. Syrjänen RK, Kilpi TM, Kaijalainen TH, Herva EE, Takala a K. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Finnish children younger than 2 years old. J Infect Dis. 2001 Aug;184(4):451–9.

38. Griffith F. The Significance of Pneumococcal Types. J Hyg (Lond) [Internet]. 1928 Jan [cited 2015 Jan 14];27(2):113–59.

39. Browall S, Norman M, Tångrot J, Galanis I, Sjöström K, Dagerhamn J, et al. Intraclonal variations among *Streptococcus pneumoniae* isolates influence the likelihood of invasive disease in children. J Infect Dis [Internet]. 2014 Feb 1 [cited 2015 Aug 14];209(3):377–88.

40. Morona JK, Miller DC, Morona R, Paton JC. The effect that mutations in the conserved capsular polysaccharide biosynthesis genes cpsA, cpsB, and cpsD have on virulence of *Streptococcus pneumoniae.* J Infect Dis [Internet]. 2004 May 15 [cited 2015 Feb 9];189(10):1905–13.

41. Weinberger DM, Trzciński K, Lu Y-J, Bogaert D, Brandes A, Galagan J, et al. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. PLoS Pathog [Internet]. 2009 Jun [cited 2015 Jan 8];5(6):e1000476.

42. Henrichsen J. Six newly recognized types of *Streptococcus pneumoniae.* J Clin Microbiol [Internet]. 1995 Oct [cited 2015 Jan 23];33(10):2759–62.

43. Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. Lancet Infectious Diseases. Elsevier; 2005. p. 83–93.

44. Johnson HL, Deloria-Knoll M, Levine OS, Stoszek SK, Freimanis Hance L, Reithinger R, et al. Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. PLoS Med [Internet]. 2010 Oct [cited 2015 Oct 29];7(10).

45. Rusen ID, Fraser-Roberts L, Slaney L, Ombette J, Lovgren M, Datta P, et al. Nasopharyngeal pneumococcal colonization among Kenyan children: antibiotic resistance, strain types and associations with human immunodeficiency virus type 1 infection. Pediatr Infect Dis J [Internet]. 1997 Jul [cited 2015 Oct 29];16(7):656–62.

46. Mbelle N, Huebner RE, Wasas AD, Kimura A, Chang I, Klugman KP. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. J Infect Dis [Internet]. 1999 Oct [cited 2015 Sep 30];180(4):1171–6.

47. von Gottberg A, Cohen C, de Gouveia L, Meiring S, Quan V, Whitelaw A, et al. Epidemiology of invasive pneumococcal disease in the pre-conjugate vaccine era: South Africa, 2003-2008. Vaccine [Internet]. 2013 Aug 28 [cited 2015 Jan 29];31(38):4200–8.

48. Feikin DR, Klugman KP. Historical changes in pneumococcal serogroup distribution: implications for the era of pneumococcal conjugate vaccines. Clin Infect Dis [Internet]. 2002 Sep 1 [cited 2015 Oct 29];35(5):547–55.

49. Harboe ZB, Benfield TL, Valentiner-Branth P, Hjuler T, Lambertsen L, Kaltoft M, et al. Temporal trends in invasive pneumococcal disease and pneumococcal serotypes over 7 decades. Clin Infect Dis [Internet]. 2010 Feb 1 [cited 2015 Sep

17];50(3):329–37.

50. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet [Internet]. 2011 Dec 3 [cited 2014 May 11];378(9807):1962–73.

51. Auranen K, Mehtälä J, Tanskanen A, S Kaltoft M. Between-strain competition in acquisition and clearance of pneumococcal carriage--epidemiologic evidence from a longitudinal study of day-care children. Am J Epidemiol [Internet]. 2010 Jan 15 [cited 2015 Sep 11];171(2):169–76.

52. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both in vitro and in vivo. Infect Immun [Internet]. 2007 Jan [cited 2015 May 26];75(1):443–51.

53. Lux T, Nuhn M, Hakenbeck R, Reichmann P. Diversity of bacteriocins and activity spectrum in *Streptococcus pneumoniae*. J Bacteriol [Internet]. 2007 Nov [cited 2015 Sep 11];189(21):7741–51.

54. Reichmann P, Hakenbeck R. Allelic variation in a peptide-inducible two-component system of *Streptococcus pneumoniae.* FEMS Microbiol Lett [Internet]. 2000 Sep 15 [cited 2015 Sep 11];190(2):231–6.

55. de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. J Bacteriol [Internet]. 2000 Sep [cited 2015 Sep 11];182(17):4696–703.

56. Pericone CD, Overweg K, Hermans PW, Weiser JN. Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. Infect Immun [Internet]. 2000 Jul [cited 2015 Sep 11];68(7):3990–7.

57. Regev-Yochay G, Trzcinski K, Thompson CM, Malley R, Lipsitch M. Interference between *Streptococcus pneumoniae* and *Staphylococcus aureus:* In vitro hydrogen peroxide-mediated killing by *Streptococcus pneumoniae.* J Bacteriol [Internet]. 2006 Jul [cited 2015 Sep 11];188(13):4996–5001.

58. Shakhnovich EA, King SJ, Weiser JN. Neuraminidase expressed by *Streptococcus pneumoniae* desialylates the lipopolysaccharide of Neisseria meningitidis and Haemophilus influenzae: a paradigm for interbacterial competition among pathogens of the human respiratory tract. Infect Immun [Internet]. 2002 Dec [cited 2015 Sep 11];70(12):7161–4.

59. Desai AP, Sharma D, Crispell EK, Baughman W, Thomas S, Tunali A, et al. Decline in Pneumococcal Nasopharyngeal Carriage of Vaccine Serotypes After the Introduction of the 13-Valent Pneumococcal Conjugate Vaccine in Children in Atlanta, Georgia. Pediatr Infect Dis J [Internet]. 2015 Jul 29 [cited 2015 Aug 3];

60. Feikin DR, Kagucia EW, Loo JD, Link-Gelles R, Puhan MA, Cherian T, et al. Serotype-specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: a pooled analysis of multiple surveillance sites. PLoS Med [Internet]. 2013 Jan [cited 2015 Apr 10];10(9):e1001517.

61. Yildirim I, Hanage WP, Lipsitch M, Shea KM, Stevenson A, Finkelstein J, et al.

Serotype specific invasive capacity and persistent reduction in invasive pneumococcal disease. Vaccine [Internet]. 2010 Dec 16 [cited 2015 Jan 23];29(2):283–8.

62. Mulholland K, Satzke C. Serotype replacement after pneumococcal vaccination. Lancet [Internet]. 2012 Apr 14 [cited 2014 May 11];379(9824):1387; author reply 1388–9.

63. Scott JR, Millar E V, Lipsitch M, Moulton LH, Weatherholtz R, Perilla MJ, et al. Impact of more than a decade of pneumococcal conjugate vaccine use on carriage and invasive potential in Native American communities. J Infect Dis [Internet]. 2012 Jan 15 [cited 2015 Apr 28];205(2):280–8.

64. Nzenze SA, Shiri T, Nunes MC, Klugman KP, Kahn K, Twine R, et al. Temporal changes in pneumococcal colonization in a rural African community with high HIV prevalence following routine infant pneumococcal immunization. Pediatr Infect Dis J [Internet]. 2013 Nov [cited 2015 Aug 14];32(11):1270–8.

65. Antonio M, Dada-Adegbola H, Biney E, Awine T, O'Callaghan J, Pfluger V, et al. Molecular epidemiology of pneumococci obtained from Gambian children aged 2-29 months with invasive pneumococcal disease during a trial of a 9-valent pneumococcal conjugate vaccine. BMC Infect Dis [Internet]. 2008 Jan [cited 2014 May 11];8:81.

66. Richter SS, Diekema DJ, Heilmann KP, Dohrn CL, Riahi F, Doern G V. Changes in pneumococcal serotypes and antimicrobial resistance after introduction of the 13-valent conjugate vaccine in the United States. Antimicrob Agents Chemother [Internet]. 2014 Nov [cited 2015 Aug 14];58(11):6484–9.

67. Cutts FT, Zaman SMA, Enwere G, Jaffar S, Levine OS, Okoko JB, 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) [Internet]. Jan [cited 2015 Sep 30];365(9465):1139–46.

68. Prymula R, Peeters P, Chrobok V, Kriz P, Novakova E, Kaliskova E, et al. Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both *Streptococcus pneumoniae* and non-typable Haemophilus influenzae: a randomised double-blind efficacy study. Lancet [Internet]. 2006 Mar 4 [cited 2015 Sep 29];367(9512):740–8.

69. Eskola J, Kilpi T, Palmu A, Jokinen J, Haapakoski J, Herva E, et al. Efficacy of a pneumococcal conjugate vaccine against acute otitis media. N Engl J Med [Internet]. 2001 Mar 8 [cited 2015 Sep 29];344(6):403–9.

70. Klugman KP, Madhi SA, Huebner RE, Kohberger R, Mbelle N, Pierce N. A trial of a 9-valent pneumococcal conjugate vaccine in children with and those without HIV infection. N Engl J Med [Internet]. 2003 Oct 2 [cited 2015 Sep 29];349(14):1341–8.

71. O'Brien KL, Moulton LH, Reid R, Weatherholtz R, Oski J, Brown L, et al. Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial. Lancet [Internet]. 2003 Aug 2 [cited 2015 Sep 29];362(9381):355–61.

72. Ruiz-Aragón J, Márquez Peláez S, Molina-Linde JM, Grande-Tejada AM. Safety and immunogenicity of 13-valent pneumococcal conjugate vaccine in infants: a meta-analysis. Vaccine [Internet]. 2013 Nov 4 [cited 2015 Aug 14];31(46):5349–58.

73. Black S, Shinefield H, Fireman B, Lewis E, Ray P, Hansen JR, 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 [Internet]. 2000 Mar [cited 2015 Sep 16];19(3):187–95.

74. Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis [Internet]. 2010 Jan 1 [cited 2015 Apr 23];201(1):32–41.

75. Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis [Internet]. 2011 Oct [cited 2015 Aug 11];11(10):760–8.

76. Millar E V, O'Brien KL, Watt JP, Bronsdon MA, Dallas J, Whitney CG, et al. Effect of community-wide conjugate pneumococcal vaccine use in infancy on nasopharyngeal carriage through 3 years of age: a cross-sectional study in a high-risk population. Clin Infect Dis [Internet]. 2006 Jul 1 [cited 2015 Sep 30];43(1):8–15.

77. Huang SS, Hinrichsen VL, Stevenson AE, Rifas-Shiman SL, Kleinman K, Pelton SI, et al. Continued impact of pneumococcal conjugate vaccine on carriage in young children. Pediatrics [Internet]. 2009 Jul [cited 2015 Aug 31];124(1):e1–11.

78. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. Thorax [Internet]. 2014 Sep 16 [cited 2015 Jan 3];

79. Madhi SA, Cohen C, von Gottberg A. Introduction of pneumococcal conjugate vaccine into the public immunization program in South Africa: translating research into policy. Vaccine [Internet]. 2012 Sep 7 [cited 2015 Feb 2];30 Suppl 3:C21–7.

80. Winn, W. C., Allen, S. D., Janda, W. M., Koneman, E. W., Procop, G. W., Schreckenberger, P. C., and Woods GL. Gram-Positive Cocci Part II: Streptococci, Enterococci, and the "Streptococcus-like" Bacteria, p. 672-764. In: Konemans's Color Atlas and Textbook of Diagnostic Microbiology. 6th ed. Wilkins LW and, editor. 672-764 p.

81. Satzke C, Turner P, Virolainen-Julkunen A, Adrian P V., Antonio M, Hare KM, et al. Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae:* updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. Vaccine [Internet]. 2013 Dec 17 [cited 2014 May 3];32(1):165–79.

82. Dube FS, Kaba M, Whittaker E, Zar HJ, Nicol MP. Detection of *Streptococcus pneumoniae* from Different Types of Nasopharyngeal Swabs in Children. PLoS One [Internet]. 2013 Jan [cited 2014 May 11];8(6):e68097.

83. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2014. Available from: http://www.r-project.org

84. Nunes MC, Jones SA, Groome MJ, Kuwanda L, Van Niekerk N, von Gottberg A, et al. Acquisition of *Streptococcus pneumoniae* in South African children vaccinated with 7-valent pneumococcal conjugate vaccine at 6, 14 and 40 weeks of age. Vaccine [Internet]. 2014 Dec 23 [cited 2015 Jan 2];

85. Ramsay JO, Giles H, Spencer G. Functional Data Analysis with R and MATLAB. 1st ed. Springer. New York: Springer-Verlag New York; 2009.

86. Kaplan SL, Barson WJ, Lin PL, Romero JR, Bradley JS, Tan TQ, et al. Early trends for invasive pneumococcal infections in children after the introduction of the 13-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J [Internet]. 2013 Mar [cited 2014 May 12];32(3):203–7.

87. Camilli R, Daprai L, Cavrini F, Lombardo D, D'Ambrosio F, Del Grosso M, et al. Pneumococcal carriage in young children one year after introduction of the 13-valent conjugate vaccine in Italy. PLoS One [Internet]. 2013 Jan [cited 2014 Jul 12];8(10):e76309.

88. Steens A, Caugant DA, Aaberge IS, Vestrheim DF. Decreased Carriage and Genetic Shifts in the *Streptococcus pneumoniae* Population After Changing the 7-Valent to the 13-Valent Pneumococcal Vaccine in Norway. Pediatr Infect Dis J [Internet]. 2015 May 26 [cited 2015 Aug 10];

89. Gounder PP, Bruce MG, Bruden DJT, Singleton RJ, Rudolph K, Hurlburt DA, et al. Effect of the 13-valent pneumococcal conjugate vaccine on nasopharyngeal colonization by *Streptococcus pneumoniae*--Alaska, 2008-2012. J Infect Dis [Internet]. 2014 Apr 15 [cited 2015 Sep 11];209(8):1251–8.

90. Ricketson LJ, Wood ML, Vanderkooi OG, MacDonald JC, Martin IE, Demczuk WHB, et al. Trends in asymptomatic nasopharyngeal colonization with *streptococcus pneumoniae* after introduction of the 13-valent pneumococcal conjugate vaccine in Calgary, Canada. Pediatr Infect Dis J [Internet]. 2014 Jul [cited 2015 Sep 11];33(7):724–30.

91. Rupa V, Isaac R, Jalagandeeswaran R, Manoharan A, Rebekah G. Epidemiology of nasopharyngeal colonization by *S. pneumoniae* in Indian infants in the first 2 years of life. Int J Pediatr Otorhinolaryngol [Internet]. 2014 Jul 25 [cited 2014 Aug 13];

92. Smith T, Lehmann D, Montgomery J, Gratten M, Riley ID, Alpers MP. Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. Epidemiol Infect [Internet]. 1993 Aug [cited 2015 Feb 16];111(1):27–39.

93. Ndlangisa KM, du Plessis M, Wolter N, de Gouveia L, Klugman KP, von Gottberg A. Population snapshot of *Streptococcus pneumoniae* causing invasive disease in South Africa prior to introduction of pneumococcal conjugate vaccines. Esposito S, editor. PLoS One [Internet]. Public Library of Science; 2014 Jan [cited 2014 Dec 12];9(9):e107666.

94. Jansen AGSC, Rodenburg GD, van der Ende A, van Alphen L, Veenhoven RH, Spanjaard L, et al. Invasive pneumococcal disease among adults: associations among serotypes, disease characteristics, and outcome. Clin Infect Dis [Internet]. 2009 Jul 15 [cited 2015 Sep 5];49(2):e23–9.

95. Lee GM, Kleinman K, Pelton SI, Hanage W, Huang SS, Lakoma M, et al. Impact of 13-Valent Pneumococcal Conjugate Vaccination on *Streptococcus pneumoniae* Carriage in Young Children in Massachusetts. J Pediatric Infect Dis Soc [Internet]. 2014 Mar [cited 2014 Jul 12];3(1):23–32.

96. Siira L, Rantala M, Jalava J, Hakanen AJ, Huovinen P, Kaijalainen T, et al. Temporal Trends of Antimicrobial Resistance and Clonality of Invasive *Streptococcus pneumoniae* Isolates in Finland, 2002 to 2006. Antimicrob Agents Chemother [Internet]. 2009 Mar 9 [cited 2015 Oct 22];53(5):2066–73.

97. Jacobs MR, Good CE, Bajaksouzian S, Windau AR. Emergence of *Streptococcus pneumoniae* Serotypes 19A, 6C, and 22F and Serogroup 15 in Cleveland, Ohio, in Relation to Introduction of the Protein-Conjugated Pneumococcal Vaccine. Clin Infect Dis [Internet]. 2008 Dec 1 [cited 2015 Oct 22];47(11):1388–95.

98. Elberse KEM, van der Heide HGJ, Witteveen S, van de Pol I, Schot CS, van der Ende A, et al. Changes in the composition of the pneumococcal population and in IPD incidence in The Netherlands after the implementation of the 7-valent pneumococcal conjugate vaccine. Vaccine [Internet]. 2012 Dec 14 [cited 2015 Sep 6];30(52):7644–51.

99. Sobanjo-ter Meulen A, Vesikari T, Malacaman EA, Shapiro SA, Dallas MJ, Hoover PA, et al. Safety, tolerability and immunogenicity of 15-valent pneumococcal conjugate vaccine in toddlers previously vaccinated with 7-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J [Internet]. 2015 Feb [cited 2015 Oct 22];34(2):186–94.

100. Zhou F, Kong F, Tong Z, Gilbert GL. Identification of less-common *Streptococcus pneumoniae* serotypes by a multiplex PCR-based reverse line blot hybridization assay. J Clin Microbiol [Internet]. 2007 Oct [cited 2015 Aug 15];45(10):3411–5.

101. Hare KM, Morris P, Smith-Vaughan H, Leach AJ. Random colony selection versus colony morphology for detection of multiple pneumococcal serotypes in nasopharyngeal swabs. Pediatr Infect Dis J [Internet]. 2008 Feb [cited 2015 Sep 11];27(2):178–80.

102. Massire C, Gertz RE, Svoboda P, Levert K, Reed MS, Pohl J, et al. Concurrent serotyping and genotyping of pneumococci by use of PCR and electrospray ionization mass spectrometry. J Clin Microbiol [Internet]. 2012 Jun [cited 2015 Jan 23];50(6):2018–25.

103. Rivera-Olivero IA, Blommaart M, Bogaert D, Hermans PWM, De Waard JH. Multiplex PCR reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuela. J Med Microbiol. 2009;58(5):584–7.

104. Steens A, Bergsaker MAR, Aaberge IS, Rønning K, Vestrheim DF. Prompt effect of replacing the 7-valent pneumococcal conjugate vaccine with the 13-valent vaccine on the epidemiology of invasive pneumococcal disease in Norway. Vaccine [Internet]. 2013 Dec 16 [cited 2014 Apr 30];31(52):6232–8.

105. Singleton RJ, Hennessy TW, Bulkow LR, Hammitt LL, Zulz T, Hurlburt DA, et al. Invasive pneumococcal disease caused by nonvaccine serotypes among alaska native children with high levels of 7-valent pneumococcal conjugate vaccine

coverage. JAMA [Internet]. 2007 Apr 25 [cited 2014 May 11];297(16):1784–92.

106. Liyanapathirana V, Nelson EAS, Ang I, Subramanian R, Ma H, Ip M. Emergence of serogroup 15 *Streptococcus pneumoniae* of diverse genetic backgrounds following the introduction of pneumococcal conjugate vaccines in Hong Kong. Diagn Microbiol Infect Dis [Internet]. 2014 Oct 2 [cited 2014 Dec 3];

107. Gill CJ, Mwanakasale V, Fox MP, Chilengi R, Tembo M, Nsofwa M, et al. Impact of human immunodeficiency virus infection on *Streptococcus pneumoniae* colonization and seroepidemiology among Zambian women. J Infect Dis [Internet]. 2008 Apr 1 [cited 2015 Sep 11];197(7):1000–5.

108. Nzenze SA, von Gottberg A, Shiri T, van Niekerk N, de Gouveia L, Violari A, et al. Temporal changes in pneumococcal colonization in HIV-infected and HIV-uninfected mother-child pairs following transitioning from 7-valent to 13-valent pneumococcal conjugate vaccine, Soweto, South Africa. J Infect Dis [Internet]. 2015 Mar 17 [cited 2015 Mar 27];

109. Nicoletti C, Brandileone MCC, Guerra MLS, Levin AS. Prevalence, serotypes, and risk factors for pneumococcal carriage among HIV-infected adults. Diagn Microbiol Infect Dis [Internet]. 2007 Mar [cited 2015 Sep 11];57(3):259–65.

110. Ramsay JO, Wickham H, Graves S, Giles H. fda: Functional Data Analysis [Internet]. Vienna, Austria: R package version 2.4.3; 2014

111. Simões AS, Tavares DA, Rolo D, Ardanuy C, Goossens H, et al. (2016) lytA-based identification methods can misidentify *Streptococcus pneumoniae*. Diagn Microbiol Infect Dis.

CHAPTER 5

Determinants Of *Streptococcus pneumoniae* Nasopharyngeal Carriage During The First Year Of Life In South African Children

5.1 INTRODUCTION

Pneumococcal conjugate vaccines (PCVs) are highly efficacious against serotype-specific invasive disease, chest radiograph–confirmed pneumonia and otitis media (OM) (1-7). PCVs have also been shown to reduce vaccine-type (VT) pneumococcal nasopharyngeal (NP) carriage among vaccinated children by reducing acquisition of VT strains (8-11). This has led to reduction in VT carriage among unvaccinated older children and adults (herd protection) largely due to reduced transmission of VT pneumococcus from vaccinated children (8,12–15). This reduction has however been offset by a parallel increase in carriage of non-vaccine-type (NVT) pneumococci through serotype replacement (16,17).

The 13-valent pneumococcal conjugate vaccine (PCV-13) has been rolled out in many sub-Saharan African countries as well as other low-middle-income countries (LMIC) where the burden of pneumococcal disease is disproportionately high (18–21). HIV/TB infections in these settings increases vulnerability to pneumococcal disease (22). Investigating the shifts in the pattern of pneumococcal colonisation provides an opportunity to assess the impact of PCV-13 and other potential risk factors for acquisition of carriage (23).

Our current understanding of the serotype-specific immunity afforded by PCVs is incomplete. In particular, the effect of prior colonisation (i.e., the pneumococcus acting as a natural immunising agent) and other multifactorial risk factors on subsequent colonisation by the same (or related) serotypes is understudied. Similarly, the effect of VT colonisation prior to a dose of PCV on vaccine efficacy has not been extensively

explored in an intensively sampled birth cohort (24–26). Continued surveillance and identification of factors influencing serotype distribution are important to fully understand the effect of vaccination on carriage and allow rational vaccine design and implementation.

5.1.1 Prevention of pneumococcal infections: Development of serotype-specific pneumococcal vaccines

5.1.1.1 Brief History

The earliest attempts in the development of pneumococcal vaccine were in 1911 when Sir Almroth Wright used whole cell pneumococcal lysates to vaccinate a cohort of South African mine workers (27, 28). The public health benefit of such a vaccine formulation was inconclusive largely because of a failure to incorporate pneumococcal serotypes known to cause invasive disease at that time, and an inadequate vaccine dose (27). The characterization of the pneumococcal capsular polysaccharide (CPS) in 1926 led to the development of the first tetravalent and hexavalent polysaccharide vaccines that were licensed in the USA in 1940 (29). However, the discovery of penicillin and other effective antimicrobial drugs for treating pneumonia, resulted in reduced interest in prophylaxis (30).

It was not until the mid-1960s that the interest in pneumococcal polysaccharide vaccines (PPVs) was revived (29). This renewed interest in prophylaxis was largely due the emergence of antibiotic resistant pneumococcal serotypes and observation that some individuals remained at risk of pneumococcal disease (29,31). A 14-valent-pneumococcal-polysaccharide vaccine (PPV-14), licensed in 1977, was replaced by a 23-

valent-pneumococcal-polysaccharide vaccine (PPV-23; Pneumovax® 23, Merck & Co., Inc.) in 1983, details on the serotypes included in each formulation are summarized in table 16 below. PPVs were primarily designed for use in older children and adults who were at risk of pneumococcal disease (32). The first pneumococcal conjugate vaccine was licensed in the year 2000.

Table 15: Comparison of serotypes included in pneumoccal vaccine formulations

Conjugate vaccines (year of licencure)						Polysaccharide vaccines	
PCV-7 (2000)	PCV-9 (2009)	PCV-10 (2009)	PCV-11 (2009)	PCV-13 (2010)	*PCV-15 (2013)	PPV-14 (1977)	PPV-23 (1983)
4	4	4	4	4	4	4	4
6B	6B	6B	6B	6B	6B	6B	6B
9V	9V	9V	9V	9V	9V	9V	9V
14	14	14	14	14	14	14	14
18C	18C	18C	18C	18C	18C	18C	18C
19F	19F	19F	19F	19F	19F	19F	19F
23F	23F	23F	23F	23F	23F	23F	23F
	1	1	1	1	1	1	1
	5	5	5	5	5		5
		7F	7F	7F	7F	7F	7F
			3	3	3	3	3
				6A	6A	6A	
				19A	19A		19A
					22F	8	8
					33F	9N	9N
						12F	12F
						20	20
						2	2
							10A
							11A
							15B
							17F
							22F
							33F
#CRM_{197}	CRM_{197}	PhiD-CV	¶TT-DT	CRM_{197}	CRM_{197}	N/A	N/A

*15-valent pneumococcal conjugate vaccine in development. #CRM_{197} = A non-toxic recombinant variant of diphtheria toxin (*Corynebacterium diphtheriae*) carrier protein. PhiD-CV = Protein carrier for the 10-valent pneumococcal conjugate vaccine that is conjugated to non-typeable *Haemophilus influenzae* protein D. ¶TT-DT = PCV11 includes serotypes 1, 4, 5,7F, 9V, 19F, 23 conjugated to a tetanus toxin while serotype 3, 14, 18C and 6B are conjugated to a diphtheria toxin.

5.1.1.2 Pneumococcal polysaccharide vaccine (PPV-23) and the unmet needs in childhood immunisation

PPVs elicit a T cell-independent immune response which is characterized by the production of serotype-specific protective antibodies (IgG, IgA and IgM are produced with the IgG2 subclass dominating the IgG response) that promote bacterial opsonophagocytosis and clearance by macrophages and neutrophils (33). Such T cell-independent antigens are poorly immunogenic in children under 2 years of age (except for serotype 3 which may induce protective antibody production) and fail to induce immune memory (34–36). In addition, PPV23 does not induce mucosal immunity, hence its failure to reduce NP carriage, an important reservoir for person-to-person transmission of pneumococci (35). Children younger than 2 years of age lack mature B cells which are necessary for T cell-dependent antibody mediated immunity (37). A vaccine is therefore effective in young children only if it can elicit a T-cell dependent serotype-specific antibody response soon after birth (38). After receiving one dose of PPV-23, infants develop antibody responses to serotypes 3, 4, 8, 9N and 18C; intermediate responses to serotypes 1, 2, 7F, 19F and 25; and poor responses to types 12, 14, 23F, and 6A and 6B, most of which are multiple drug resistant (15,39). The antibody titers decrease rapidly within a few months of immunization in younger children and a second dose fails to induce an anamnestic booster response (40,41). The immunological basis for the poor immunogenicity is largely due to the inability of most polysaccharide antigens to associate with major histocompatibility complex (MHC) class II molecules on the surface of antigen presenting cells (APCs; includes dendritic cells, macrophages, B cells). Consequently, B cells that recognize polysaccharide antigens using their surface expressed IgM cannot get help from T cells (42). The failure of PPVs

to induce T cell help results in the poor persistence of vaccine induced antibody titers and the requirement for re-vaccination (37).

5.1.1.3 Pneumococcal conjugate vaccine: A tale of two paradigms

The failure of PPVs to induce a T cell-dependent immune response and the added costs incurred in re-vaccination stimulated the development of protein based PCVs (4). The 7-valent pneumococcal conjugate vaccine (PCV-7; Prevenar®, Pfizer) contains capsular polysaccharides from pediatric serotypes 4, 6B, 9V, 14, 18C, 19F and 23F, which are responsible for 80% of invasive pneumococcal diseases in developed countries (5, 6, 10, 27). The 10 and 13-valent pneumococcal conjugate vaccines (PCV-10, -13; Prevenar®, Pfizer), currently incorporated in routine immunisation programs in many countries contain all serotypes included in PCV-7 with two (1, 5) and five (1, 5, 7F, 6A and 19A) additional serotypes in the case of PCV-9 and PCV-13, respectively. Altogether, these serotypes account for 65-80% of the serotypes associated with pneumococcal disease among children from developed countries (25). T-cell dependent conjugate vaccines not only elicit protection against systemic bacteremia through IgG but also induce protection against colonisation of the host mucosal surfaces through production of IgA which is secreted on the mucosal surface and consequently reduces NP carriage of pneumococcal serotypes included in the PCV formulation (42–44).

5.1.1.4 Pneumococcal conjugate vaccines mode of action

PCVs contain purified capsular polysaccharide covalently linked to a protein carrier which are taken up by surface expressed antibodies on B cells, internalized, co-processed with the carrier protein and then presented to T cells (42). The peptide (protein carrier) anchors the polysaccharide epitope to the MHC and allows its

presentation to polysaccharide-specific CD4+ helper-T-cells. The interaction between the polysaccharide epitope and the surface expressed T cell receptor (TCR) triggers the production of cytokines (interleukin (IL)-4, IL-5 and IL-6) responsible for B cell antibody class switching (43). These cytokines stimulate B cells to proliferate, class-switch from IgM to IgG antibody production, initiate affinity maturation and establish immune memory (24). Immune memory primes the immune system for future natural exposure or for a subsequent booster effect with re-vaccination, figure 28.

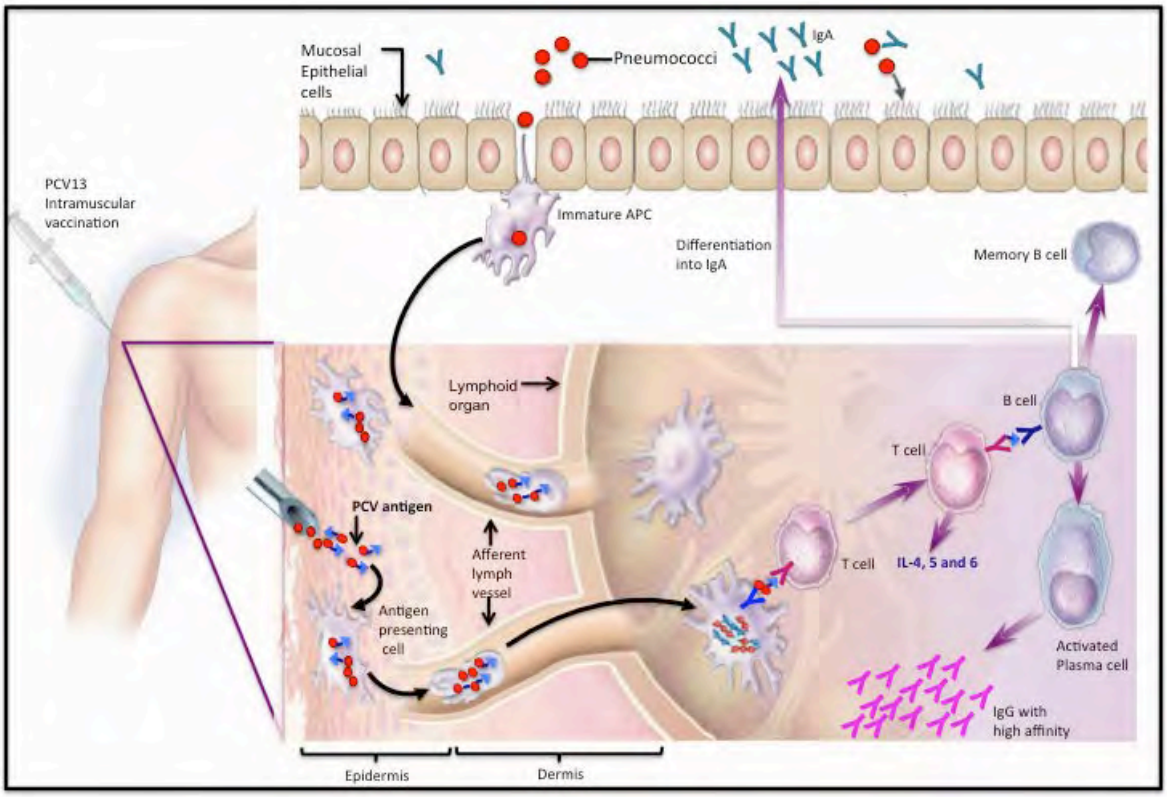

Figure 26: Antibody response to pneumococcal polysaccharide conjugate vaccines (PCV). After intramuscular injection with PCV, the pneumococcal polysaccharide (Red circles) linked to the carrier peptide (Blue triangles) is taken up by Antigen Presenting Cells (APC, which include Macrophages, Dendritic and B cells), internalised and processed. The peptide then anchors the polysaccharide epitope into the Major Histocompatibility complex Class II (MHC) cavity, this allows the presentation of the polysaccharide epitope to polysaccharide-specific type 2 helper T (Th2) cells. The activated Th2 cells secrete interleukin (IL) 4, 5 and 6 cytokines which causes B cells to proliferate, switch from producing immunoglobulin M (IgM) to produce IgG and establish immunological memory. These cells mature in the lymphoid follicles; only cells that express very high affinity IgG become plasma B cells and secrete IgG that binds to specific

pneumococcal capsule and mediates opsonic activity and complement-mediated bactericidal activities, modified from Pletz *et al.* (43).

5.1.1.4 Impact of pneumococcal conjugate vaccines on nasopharyngeal carriage

The mechanisms through which PCVs directly protect against VT carriage acquisition are incompletely understood. It has however been suggested that this protection is mediated by secretory IgA rather than IgG. High titers of naturally acquired or vaccine-induced circulating serotype-specific antibody correlate with protection against acquisition(12,45). Healthy children previously exposed to pneumococci showed an increase in type-specific secretory IgA concentration in saliva, suggesting that carriage induces production of mucosal antibodies (46). In addition, toddlers in Israel (47) and adults in the United Kingdom (45) showed increases in levels of IgG specific to carried serotypes such as 14 and 23F, and this rise was associated with reduced odds of reacquiring the same serotype.

The success of PCVs in preventing serotype-specific invasive disease has been tempered by concern over replacement of VT serotypes with NVT serotypes in carriage and invasive clinical isolates (16,17,23,48,49). Previous studies have shown a decrease in NP carriage of VT serotypes following introduction of PCV immunisation (6)(50)(51)(52). However, other studies evaluated the effect of PCV-9 on carriage and did not find any significant differences in the rates of VT carriage between the vaccine group and the control arm (14,18). Some initial studies amongst Gambian infants evaluated the effect of administering two to three doses of a pentavalent pneumococcal conjugate vaccine (PCV-5) followed by PPV-23 booster dose (19). Twenty months after primary immunization, there was a reduction in carriage of VT serotypes in children

who had received PCV-5 in infancy and PPV23 at age 18 months. However, there was no difference in all vaccinated groups with regard to overall carriage (19) as children who received at least three doses of PCV carried significantly more NVT pneumococci than children in the unvaccinated group (77% *vs.* 43%, $p = 0.002$).

5.1.2 Prevention of pneumococcal infections: protection induced by prior NP colonisation

5.1.2.1 Pneumococcal colonisation as a natural immunizing event

The establishment of both antibody mediated and antigen-specific T cell immune responses following pneumococcal colonisation have been shown to protect against subsequent colonisation and progression to invasive disease in murine models (54). Similarly, prior pneumococcal colonisation, in both infant and adult human subjects, can in itself act as an immunizing event (53,54). This protection has been associated with an increase in serum antibody titers against the pneumococcal capsule and surface expressed proteins. Ferriera *et al* recently showed the direct evidence of this protection using a controlled experimental human model of pneumococcal colonisation to determine the immunizing effects of prior NP colonisation by pneumococcus. In this study, healthy adult volunteers aged 18-60 years were inoculated with pneumococcal serotype 6B. The authors reported successful colonisation in 41% (29/70) of healthy volunteers. Similarly, exposure to pneumococcal antigens (specifically protein and not capsular) without carriage can also protect against subsequent carriage acquisition (55). This protection has been associated with increases in the serum and mucosal antibody titres against pneumococcal surface proteins (*psp*A). Antigen-specific T cells

play an important role in the generation of optimal antibody titers and immune memory (56).

5.1.2.2 Impact of prior NP carriage on vaccine efficacy

Children who were colonised by pneumococcus in early life were more likely to reacquire pneumococcus which correlates with increased risk of transmission (57). Immunization with PPV-23 results in with serotype-specific hyporesponsiveness, that is, individuals initially primed with PPV-23 tend to have lower functional anti-capsular antibody titers after challenge with a PPV-23 booster (58). This phenomenon has also been shown during invasive pneumococcal disease (IPD) in which it was speculated that the observed immune paralysis was due to large pneumococcal polysaccharide antigen loads during IPD (59). Pneumococcal conjugate vaccines (PCVs) were initially shown to overcome the hyporesponsiveness induced by PPV-23 (60–62). However, recent studies have shown that NP colonisation prior to the first PCV dose in young infants could also induce hyporesponsiveness to the specific colonizing serotypes (63-65). Madhi *et al* showed that South African children colonised by VT (6B, 19F and 23F) pneumococci prior to completion of a primary series of PCV-7 immunization had inferior antibody responses to the colonising serotype (65).

All the studies described above were conducted prior to the introduction of PCV-13. There are limited data to demonstrate whether prior colonisation by VT pneumococcus at, or immediately before the first PCV-13 dose, affects the serotype-specific carriage acquisition in response to primary PCV-13 immunization series and/or to the booster PCV-13 dose.

5.1.3 Risk factors associated with increased acquisition of pneumococcal carriage

Several environmental and host factors have been implicated in increased pneumococcal carriage. Daycare attendance and overcrowded family homes are a potential source for effective transmission of pneumococci. Children in crowded family homes often harbor and transmit pneumococci to other family members. The high rates of antibiotic prescription among children attending daycare centres select for resistant strains which disseminate among contacts. A large-scale longitudinal carriage study conducted in Israeli day-care center children showed that 56% (334/594) of pneumococcal isolates were resistant to at least one antibiotic (63). Antibiotic use and PCV vaccines both induce selective pressure that seems to be driving pneumococcal capsular replacement (64).

NP carriage prevalence and acquisition rates have been shown to vary considerably by season, with peaks in winter months for temperate climates (13,65,66) and during the rainy season for Sub-Saharan African (67). These weather patterns favor indoor living leading to temporary crowding in poorly ventilated structures which may enhance person-to-person transmission (67). However, a study in Finland showed a fluctuation in prevalence throughout the year, as opposed to a particular seasonal pattern (68).

The incidence of pneumococcal disease is not only a function of the frequency of carriage but may also be related to differences in the invasive disease potential of particular pneumococcal strains (69-71) and the prevalence of preceding viral infections (72-75). Other host risk factors that may predispose an individual to pneumococcal carriage and disease include age, active/passive smoking, inadequate breast feeding, asthma, congestive heart failure, diabetes mellitus and individuals with

primary or acquired immunodeficiencies (76,77). Comparisons of seasonal variations and other environmental determinants provide an opportunity to assess drivers of pneumococcal carriage prevalence and acquisition rates.

5.2 Aims and Objectives

The overall aim of the analysis described in this chapter is to identify factors associated with pneumococcal carriage in a South African birth cohort. These include:

1. What is the impact of each successive dose of PCV-13 on acquisition of carriage by VT and NVT serotypes at the level of the individual child?
2. Does prior (early life) carriage with VT or NVT serotypes influence the efficacy of PCV-13 in preventing carriage with the same/related serotypes?
3. Which risk factors (demographic, socio-economic and environmental) are associated with acquisition of pneumococcal carriage?

5.3 MATERIALS AND METHODS

5.3.1 Study setting

A detailed description of the study setting is given in chapter 4 (4.4.1).

5.3.2 Participant enrollment

A detailed account of participant enrolment and retention is given in section 4.4.2 of chapter 4. Detailed assessment of risk factors was conducted by trained study nurses during each scheduled visit at 6, 10 and 14 weeks and at 6, 9 and 12 months. These interviews were conducted in the mother's preferred language (English, Afrikaans or

Xhosa). The use of household biofuels (wood, coal or kerosene), presence of household smokers and the number of people sharing the bedroom with the child (overcrowding), history of respiratory symptoms were established. Information on daycare attendance was also collected. In addition, questionnaires about maternal health were administered antenatally. Obstetric care was provided at Paarl Hospital where all births took place. The mother's HIV status was also established. If the mother was HIV positive, the latest available CD4 count was recorded. Children were classified as HIV exposed if the mother was HIV infected at the time of the pregnancy but the child tested HIV negative. Seasons of the year were categorized as Summer (December–February), Autumn (March–May), Winter (June–August) and Spring (September–November).

5.3.3 Collection, processing and pneumococcal capsular typing

Bi-weekly NP swabs were collected from a subset of 137 children enrolled in the DCHS over the first year of life, as described in chapter 3 (Section 3.3).

5.3.4 Statistical analysis

Initial exploratory statistics were performed using STATA software (Stata Corporation, College Station, TX) and the rest of the analysis were performed using the openly available statistical environment R, version 3.1.1(78). Functional data analysis (FDA) including curve registration computations was performed using facilities provided by the University of Cape Town's ICTS High Performance Computing system

(HPC; http://hpc.uct.ac.za). Chi square test was used to assess the association between carriage prevalence and seasonal variation.

5.3.4.1 Functional Data analysis methodology

One of the limitations when analyzing the effect of sequential PCV-13 vaccination is correcting for the variations between children in the actual ages at which the vaccine was administered. To overcome this, FDA was used (specifically curve registration) to 'align' the vaccination points in order to compare the serotype-specific colonisation patterns in response to vaccinations. This analysis was performed as follows:

In FDA a continuous underlying function is assumed. For our data, the presence and absence of each serotype is observed at two-weekly intervals. The underlying function has values zero when the serotype is absent and one when the serotype is present. Hastie *et al.* (79) discuss nonparametric logistic regression where spline functions are used to model presence-absence data. With the R package fda (78) cubic bsplines are used in a logistic regression model to fit a function in continuous time to the observed values from every two weeks.

In logistic regression the log-odds is modelled as a linear function of time:

$$log\left(\frac{P(serotype\ present)}{P(serotype\ absent)}\right) = f(t) + \varepsilon$$

where $f(t)$ is a cubic bspline function of time and ε accounts for random noise. From the equation above it follows that the estimated $P(serotype\ present)$ can be written as

$$estimated\ P(serotype\ present) = \frac{\exp\left(f(t)\right)}{1 + \exp\left(f(t)\right)}$$

In Figure 29, the observed data is represented by black dots. The serotype is absent at all observed time points except weeks 32, 36 and 42. The blue line provides the estimated probability that the serotype is present as a smooth function over time.

The recommended vaccination time point is at 6 weeks, 14 weeks and 40 weeks. Not all infants however received the vaccinations at exactly these time points. For the infant used in the example in Figure 29, the vaccinations took place at 6, 15 and 46 weeks, figure 30 below.

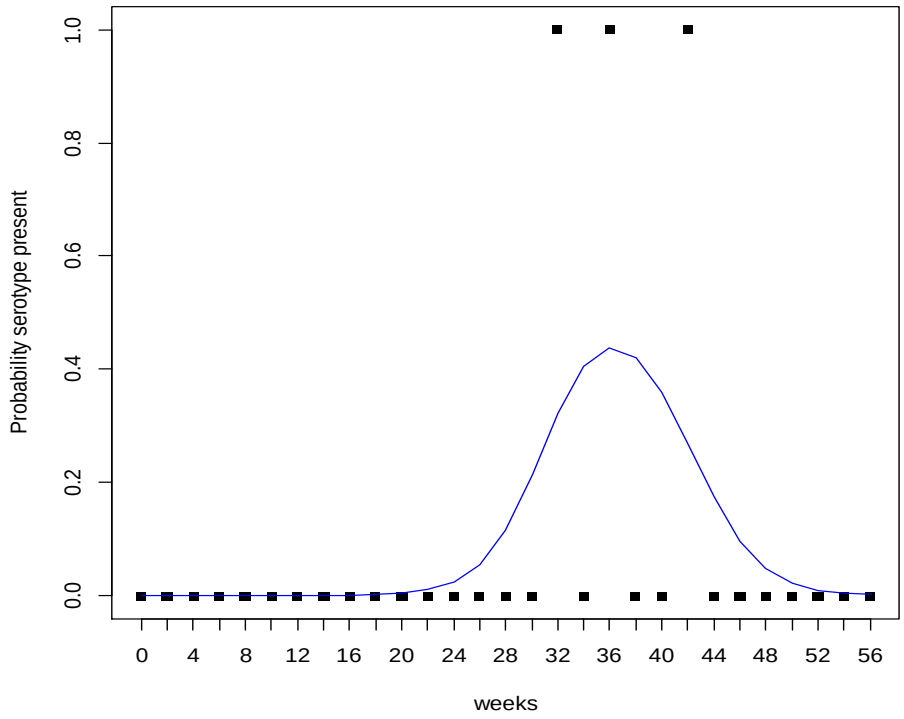

Figure 27: Observed data and non-parametric logistic regression fit of a single serotype in one infant.

Functional data analysis therefore allows for so-called curve registration. The method of landmark registration uses a time warping function to stretch and/or shrink time so that certain landmarks are aligned, figure 30.

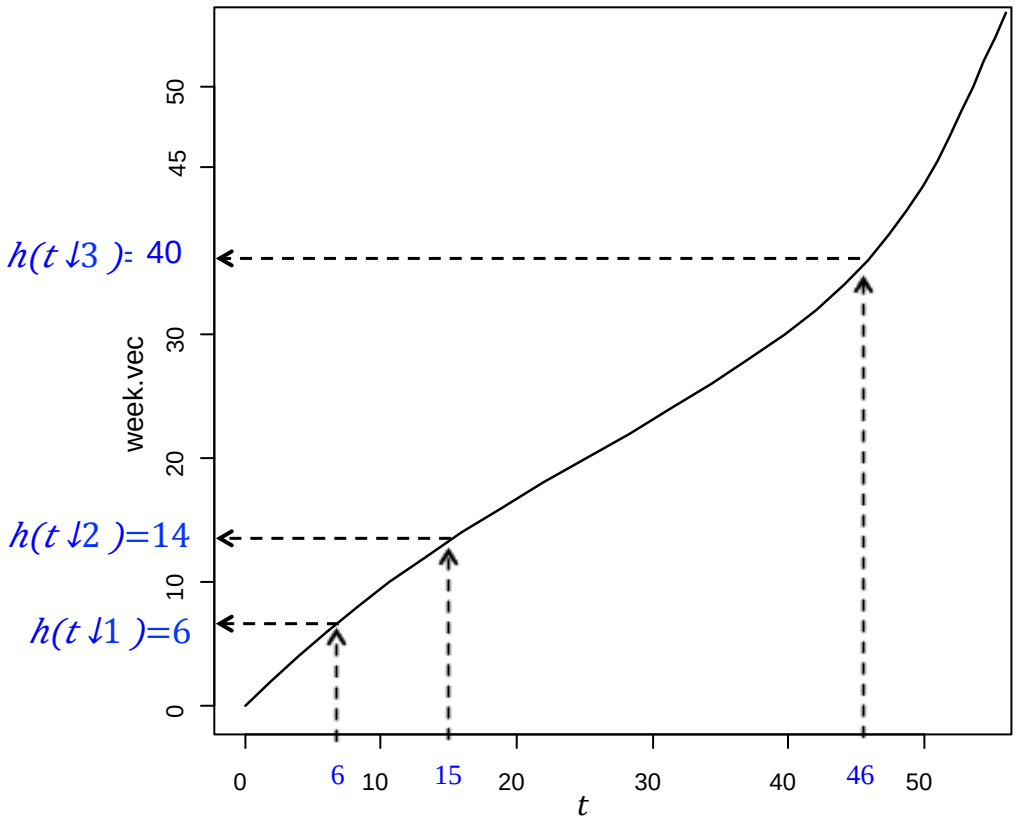

Figure 28: Time warping function for the curve registration of figure 2.

Appling a time warping function is ideal for investigating the effect of vaccination on the nonparametric logistic regression fits. For the infant above, the first vaccination took place at the recommended time point of 6 weeks. Since the second vaccination was slightly late, at 15 weeks, the effect of the vaccination will only be visible after 15 weeks. Comparing different infants with different vaccination time points will make it very difficult to judge the effect of the vaccination. By using landmark registration, the time axis between 6 and 15 weeks is slightly shrinked, so that the nonparametric logistic regression curve value at the time of the second vaccination (15 weeks) appears in line with 14 weeks, figures 30 above. Once each individual infant's estimated functions are aligned in this way, the effect of vaccinations at the landmark time points of 6, 14 and 40 weeks can be evaluated as if all infants received their vaccinations at exactly these time

points.

A different time warping function is needed for each infant, depending on when the infant received the three vaccinations.

Once the vaccination time points for all infants have been aligned with the landmarks using the R function landmarkreg in the FDA package, the shapes of the curves can be compared. A mean curve can be constructed, representing the **mean estimated probability that the serotype is present at each time point**.

5.3.4.2 Anderson & Gill Model

The *Anderson – Gill (AG) model* was employed to assess the independent association of risk factors and pneumococcal carriage. This model extends the standard Cox proportional Hazard Model (CPH model, discussed in chapter 4) into a counting process and then further considers repeated events by clustering individuals (80). The model however assumes independence between repeated events per individual and it uses a common baseline hazard function for all events and estimates a vector of common parameters for the vector of covariates regardless of the order in which the events occurred.

5.3.5 Ethical considerations

The study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town (HREC ref: 401/2009 and 740/2013) and the Western Cape Provincial Child Health Research Committee. Mothers provided written informed consent at enrolment and annually thereafter.

5.4 RESULTS

5.4.1 Incidence of pneumococcal carriage acquisition

Differences in baseline characteristics of children between the two enrolment sites, TC Newman and Mbekweni were recorded, Table 17. All the children enrolled from TC Newman were of mixed race origin (coloured African) whereas children from Mbekweni were largely ethnic black Africans with exception of one child who was of mixed race origin. Overall, 24% (33/137) of the mothers were HIV infected, with higher prevalence at Mbekweni (39% vs. 5%, $p < 0.0001$), table 17. Of all the children born to HIV positive mothers at Mbekweni, only one child was confirmed to be HIV positive. The rest of the children were HIV exposed but uninfected. None of the children born to the three HIV positive mothers at TC Newman were HIV infected. Daycare attendance was 20% and 25% of children at TC Newman and Mbekweni. Exposure to tobacco smoke (self-reported smoking from the mothers) was more common at TC Newman than Mbekweni, 48% (29/60) vs 9% (7/77), table 17.

Table 16: Baseline characteristics of children included in the present study

	TC Newman (n= 60)	**Mbekweni (n =77)**	***P* - value**
Gender			
Female	34 (57%)	44 (57%)	1.00
Male	28 (43%)	33 (43%)	1.00
Race and ethnic origin			
Black African	0	76 (99%)	**<0.0001**
Mixed Race	60 (100%)	1 (1%)	**<0.0001**
Preterm			
No	51 (85%)	66 (86%)	0.86
Yes	9 (15%)	11 (14%)	0.86
Mode of delivery			
Normal vaginal	45 (75%)	63 (82%)	0.32
Normal vacuum	1 (2%)	0	0.21
Elective caesarean	4 (6%)	6 (8%)	0.65
Emergency caesarean	10 (17%)	8 (10%)	0.22
HIV exposure			

No	57 (95%)	47 (61%)	**<0.0001**
Yes	3 (5%)	30 (39%)*	**<0.0001**
Feed choice			
Exclusive breastfed	7 (12%)	15 (20%)	0.21
Mixed fed	28 (47%)	44 (57%)	0.25
Never breastfed	25 (42%)	18 (23%)	**0.02**
Other Children			
No	42 (70%)	48 (62%)	0.33
Yes	17 (28%)	29 (38%)	0.22
Low birth weight			
No	48 (80%)	71 (92%)	**0.04**
Yes	12 (20%)	6 (8%)	**0.04**
Daycare[1]			
No	47 (78%)	58 (75%)	0.68
Yes	12 (20%)	19 (25%)	0.49
Smoking status of Mother			
No	24 (40%)	60 (78%)	**<0.0001**
Yes	29 (48%)	7 (9%)	**<0.0001**
Missing data	7 (12%)	10 (13%)	0.86
Household crowding			
Less than three people	26 (43%)	46 (60%)	**0.04**
Four or five people	23 (38%)	15 (19%)	**0.01**
More than five	7 (11%)	3 (4%)	0.11
Missing data	4 (7%)	13 (17%)	0.08

*Only one child born to an HIV mother was HIV positive. [1]No information on daycare attendance was available for one child.

Children enrolled in the present study (n=137) accrued 48826 child-days of follow-up. The lowest incidence of pneumococcal acquisition was observed during the first 2 weeks of follow-up, 0.7 episodes per 100 child-days whereas the highest incidence was at 21 weeks, 2.2 episodes per 100 child-days, figure 31.

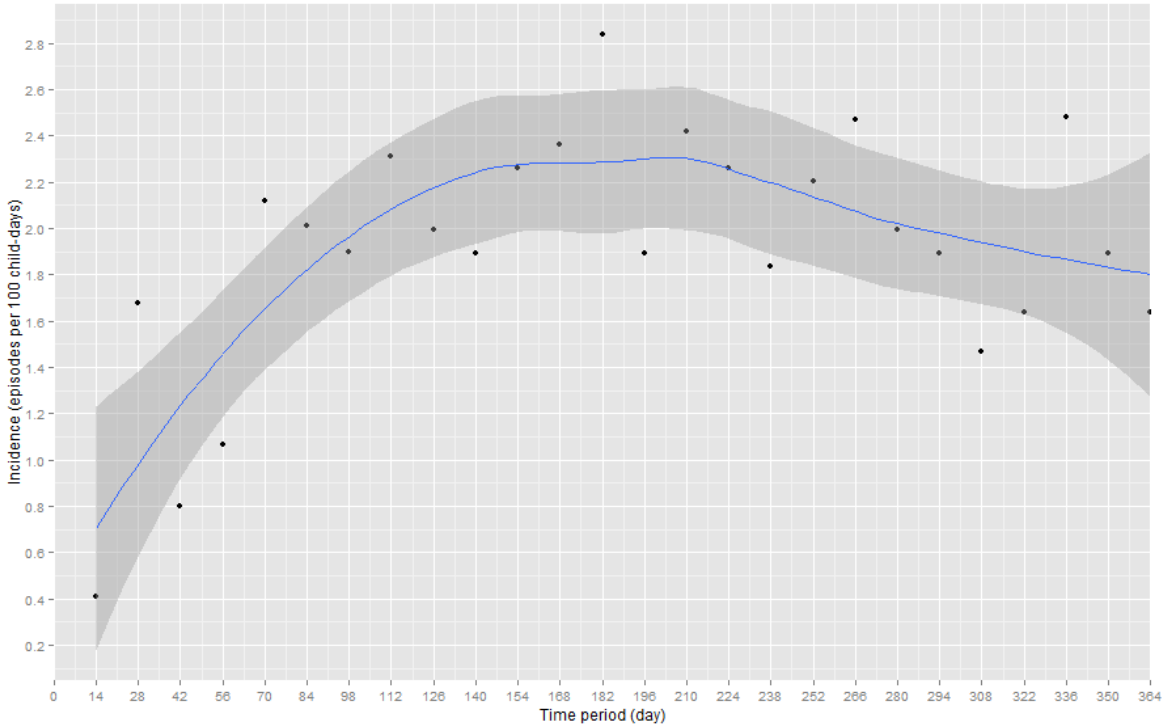

Figure 29: Incidence of pneumococcal acquisition over the first year of life. Black dots show carriage incidence. Blue line shows fitted values. Upper and low bound grey areas represents the 95% confidence interval.

5.4.2 PCV-13 vaccine coverage

All except three children received their complete schedule of PCV-13 vaccine doses. These three children received their 1st and 2nd doses but did not receive their 3rd dose of PCV-13. The median time to the 1st dose of PCV-13 was 6.1 weeks (Interquartile range, IQR = 5.1 – 13.4 weeks), while the median time to the 2nd and 3rd dose was 15.0 weeks (IQR = 10.0 – 28.9 weeks) and 39.3 weeks (IQR = 29.1 – 64.0 weeks) respectively, figure 32. The first (6 weeks) and second (14 weeks) primary PCV-13 doses were given within 2 weeks of the scheduled time in 94% (129/137) and 72% (99/137) of the children respectively; conversely, 6% (8/137) and 28% (38/137) of the infants had a more than 2 weeks delay in receiving their scheduled first and second PCV-13 doses respectively.

The booster dose at 40 weeks was given on time in 89% (122/137) with only 11% (15/137) of the children delayed in receiving their booster dose.

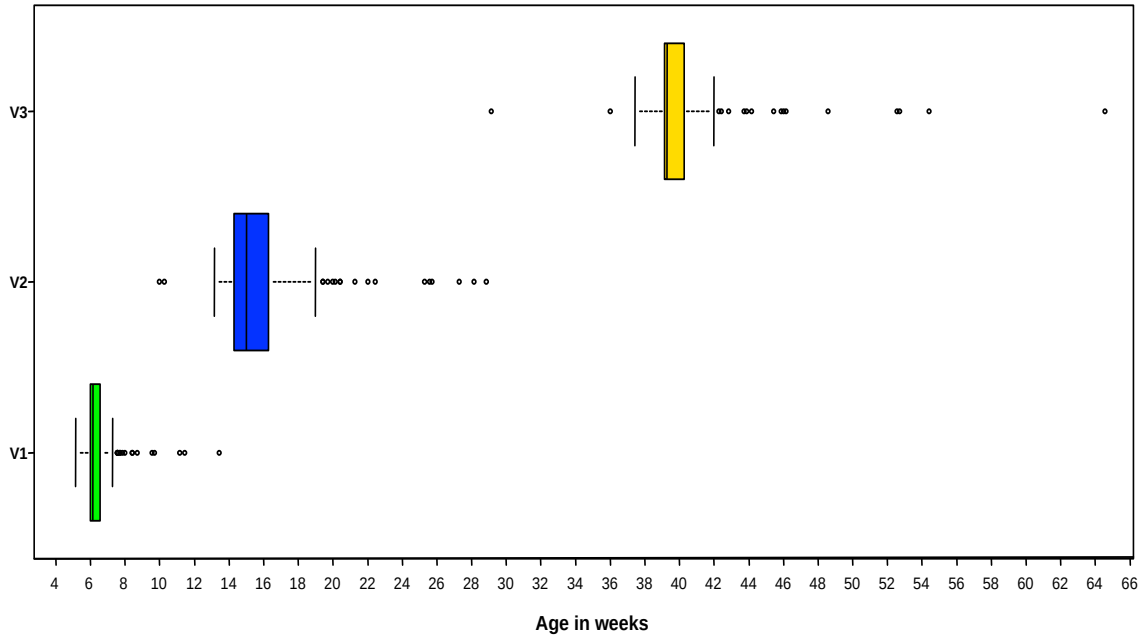

Figure 30: Box and whisker plot representing the median age at PCV-13 immunisation. The outer bounds of each box (green, blue and yellow) represent the interquartile ranges (IQR; difference between 25th and 75th percentile), the median is indicated with a vertical line inside the box, whereas the vertical lines extending from the box ("whiskers") indicate variability in the data. Outliers are shown as individual points. V1 (green plot) = 1st dose of PCV-13, V2 (blue plot) = 2nd dose of PCV-13 and V3 (yellow plot) = 3rd dose of PCV-13 administration.

5.4.3 Impact of sequential PCV-13 doses on mucosal carriage of VT and NVT serotypes

After correcting for the differences in the vaccination dates (figure 32 above) through FDA and curve registration, the mean probability of VT serotypes carriage remained approximately constant over the first year of life with no significant increments or reductions after each sequential dose of PCV-13 immunisation, figure 33. There were gradual increments in the mean probability of NVT serotypes carriage through the first year of life with a peak at 24 weeks before reaching a plateau, figure 33, whilst

probability of carriage of VT pneumococci remained more or less constant after the first month of life.

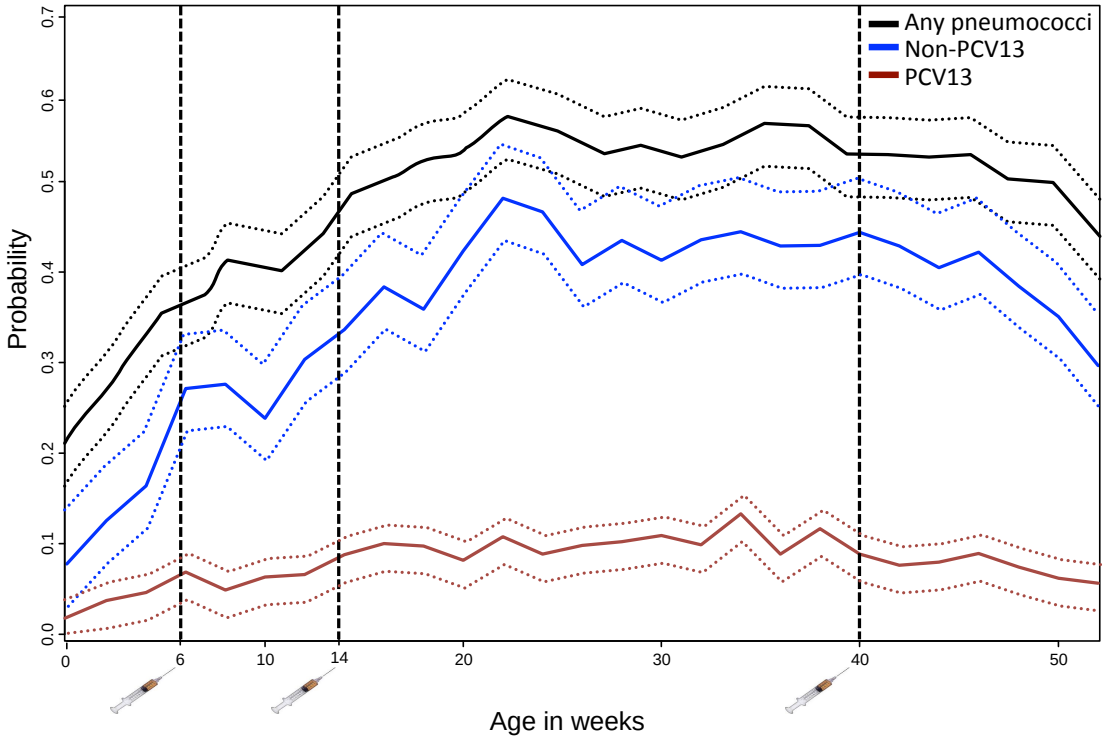

Figure 31: Nasopharyngeal carriage of pneumococci in relation to PCV-13 vaccine doses. Black line = mean probability of pneumococcal carriage regardless of the serotype. Red line = mean probability of PCV-13 vaccine serotype carriage. Blue line = Mean probability of non-PCV-13 serotypes carriage. The dotted lines represent the upper and lower bound 95% confidence interval while the vertical dashed lines indicate the PCV-13 immunisation dates.

This analysis was repeated for individual VT serotypes 19A, 19F, 6A and 9V, figure 34 below (additional mean registered curves for 6B, 14, 3 and 4 are shown in appendix) Serotype 1 was not included in the plots below because only one child carried this particular serotype. Although serotypes 6A and 9V slightly increased in proportion following the 2nd of PCV-13, the proportions reverted back to the initial levels after the 3rd dose of PCV-13. . There was no clear relationship between carriage of each individual VT serotype and each vaccine dose.

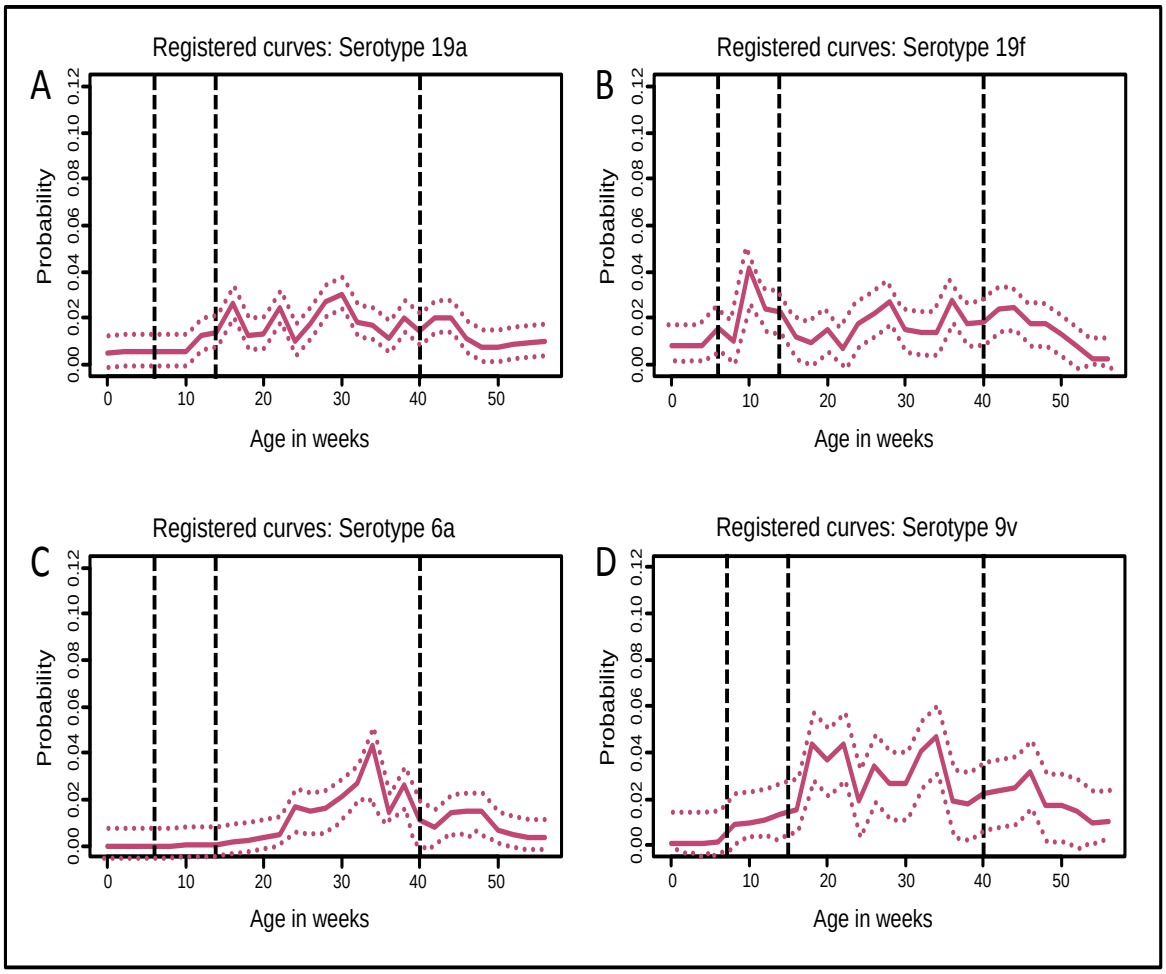

Figure 32: Mean registered curves showing the mucosal carriage of individual PCV-13 vaccine serotypes with 95% upper and lower confidence intervals (dotted red lines). A = Serotype 19A, B = Serotype 19F, C = Serotype 6A and D = Serotype 9V. The dotted lines represent the upper and lower bound 95% confidence interval while the vertical dashed lines indicate the PCV-13 immunisation dates.

The individual NVT registration curves for six (6) most common NVT are shown in are shown below, figure 35. Serotype 15B and 15C show a steep increase after the sixth week of life, plateau after the 14th week and drop after the 40th week.

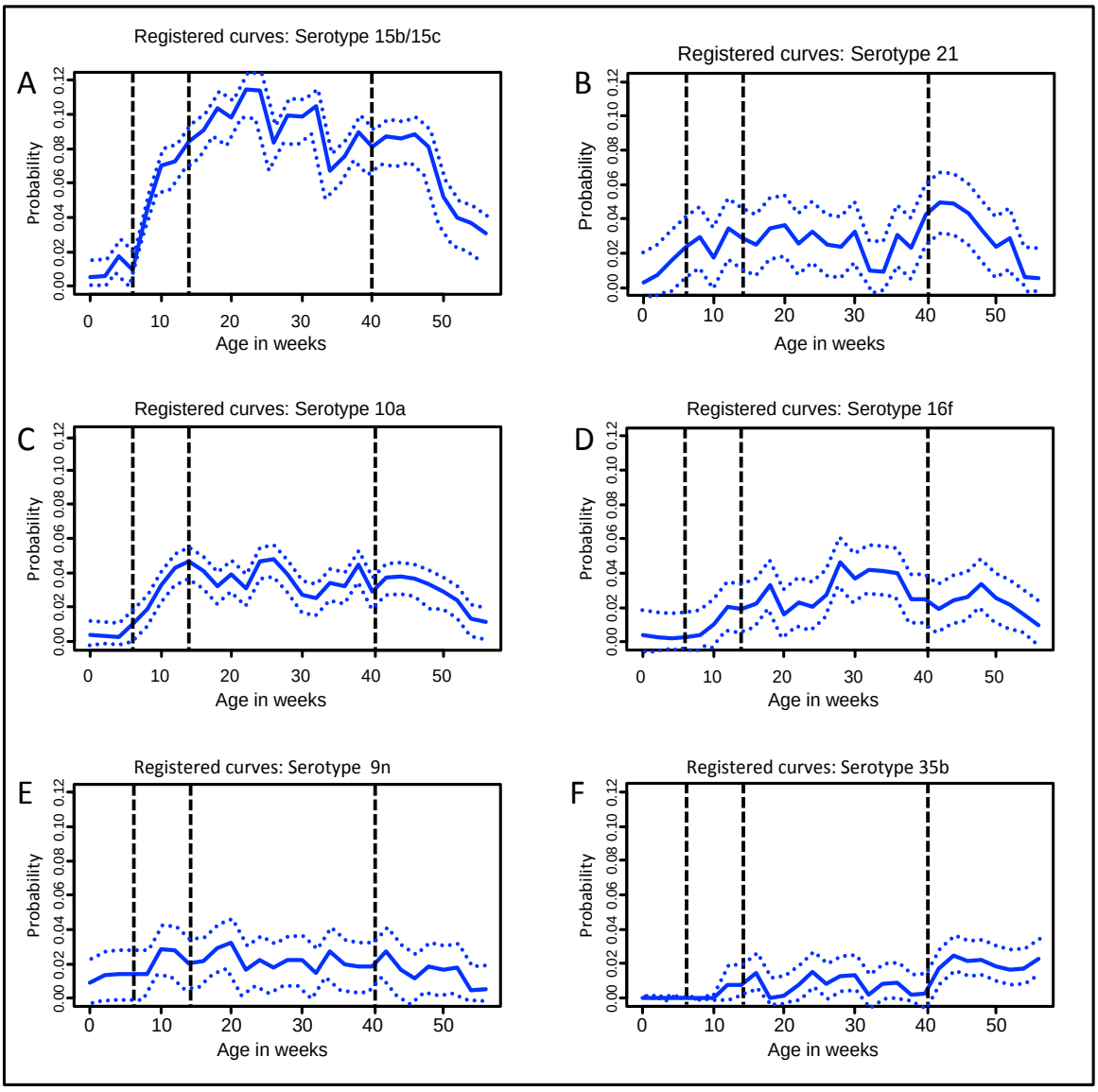

Figure 33: Mean registered curves of six (6) commonest non-PCV-13 vaccine serotypes with 95% upper and lower confidence intervals (dotted red lines). A= Serotype 15B/15C, B= Serotype 21, C= Serotype 10A, D= Serotype 16F, E= Serotype 9N and F= Serotype 35B. The dotted lines represent the upper and lower bound 95% confidence interval while the vertical dashed lines indicate the PCV-13 immunisation dates.

5.4.4 Mucosal carriage of VT serotypes stratified by HIV exposure.

VT carriage as seen in figure 34 above was further stratified according to the children's HIV exposure, corrected for site, figure 36 below. The proportion of VT carriage was

generally higher among HIV unexposed children compared to HIV exposed children although this difference was not statistically significant.

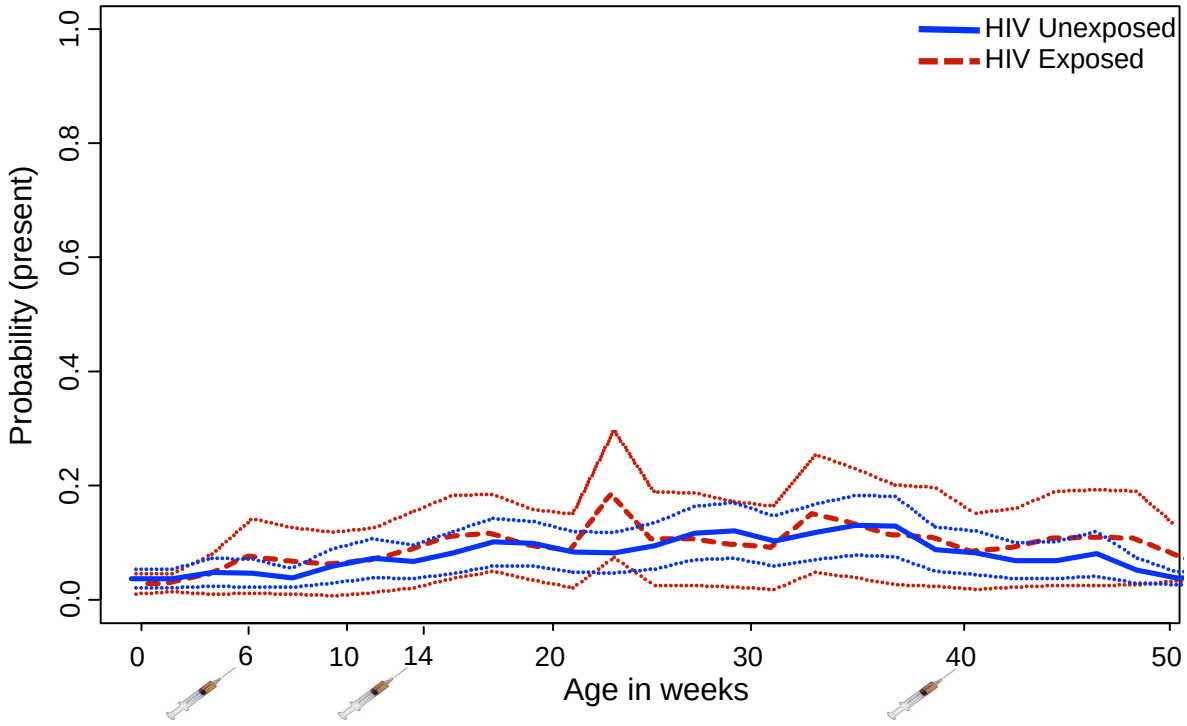

Figure 34: Mucosal carriage of PCV-13 serotypes stratified by HIV exposure. There was no significant difference in VT carriage amongst children who were HIV unexposed, solid blue line, and HIV exposed children, thick dotted red line, thinner dotted lines represent the 95% CI lines.

5.4.5 Mucosal carriage of NVT serotypes stratified by HIV exposure.

Although the prevalence of NVT carriage was higher than VT serotypes, there was equally no difference in the mean probability of NVT carriage was significantly higher among HIV unexposed children compared to HIV exposed children after stratifying for HIV exposure, figure 37.

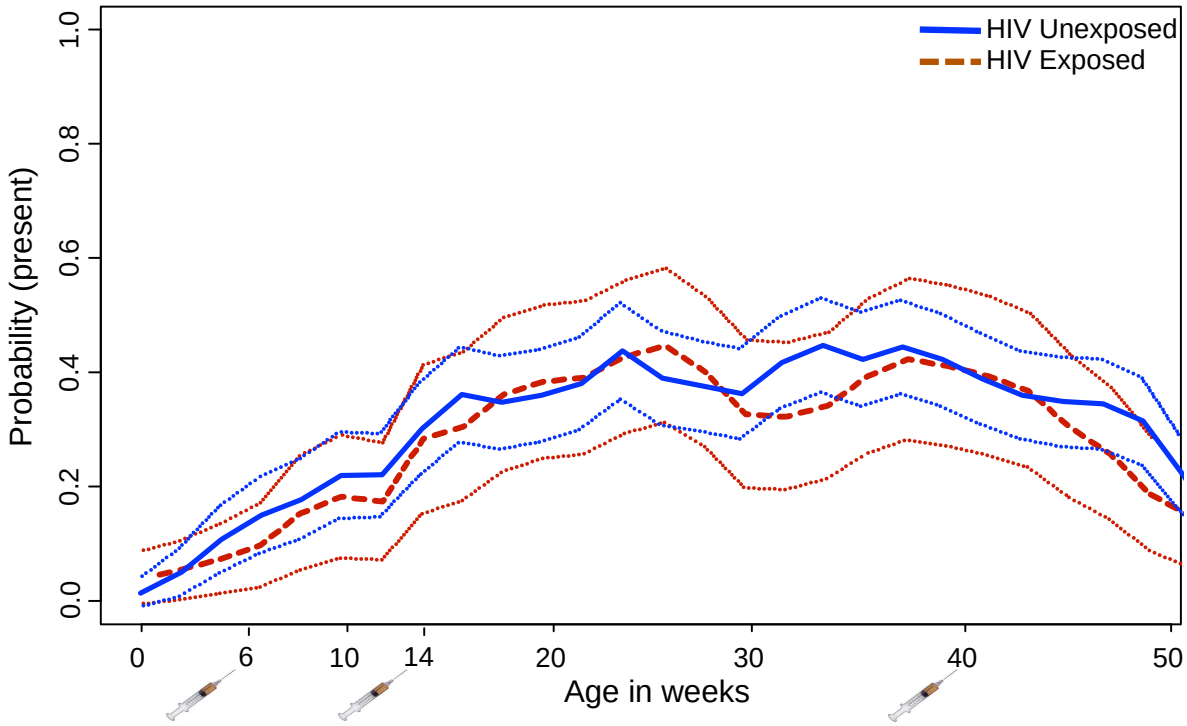

Figure 35: Mucosal carriage of non-PCV-13 serotypes stratified by HIV exposure. There was no significant difference in NVT carriage amongst children who were HIV unexposed, solid blue line, and HIV exposed children, thick dotted red line, thinner dotted lines represent the 95% CI lines

5.4.6 Impact of prior colonisation on vaccine efficacy in preventing carriage

To determine whether prior (early life) mucosal carriage with VT or NVT serotypes influences the efficacy of PCV-13 in preventing mucosal carriage with the same/related serotypes, the proportion of children colonised with pneumococci prior to any dose of PCV-13 was compared to the proportion of children colonised after sequential administration of PCV-13, Figure 38.

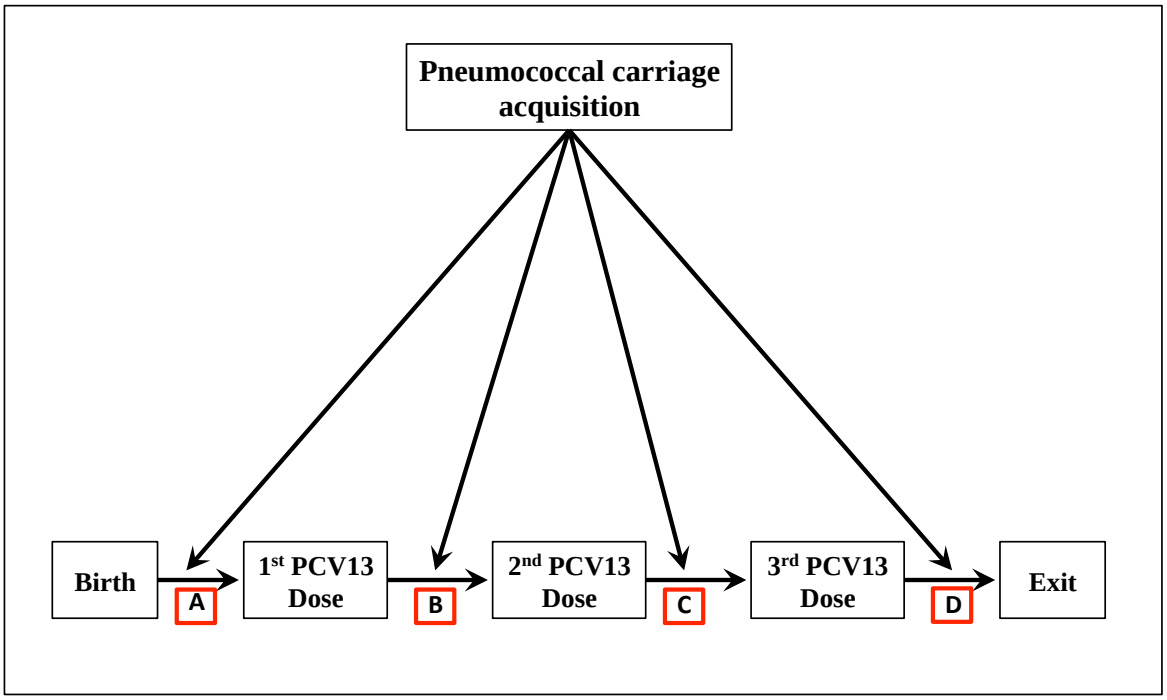

Figure 36: Approach to assess the differences in the proportion of children colonised. The proportion of children colonised prior to any PCV-13 dose was compared to the proportion of children colonised after each dose of PCV-13 vaccine. For convenience, four intervals are described (Before 1st dose, A; between the 1st and 2nd dose, B; between 2nd and third dose, C and after the 3rd dose, D;).

5.4.6.1 Does the proportion of VT and NVT significantly differ between sampling interval?

A comparison of the proportion of VT pneumococci to NVT pneumococci in each interval (A-D), Figure 13 below, showed that there was no significant difference in the proportion of VT and NVT in all the intervals considered (A-D) ($p > 0.05$). This is consistent with the overall trend in the point prevalence observed in this cohort, figure 21, chapter 4. At all sampling points considered, the ratio of VT to NVT is approximately 20:80, figure 39.

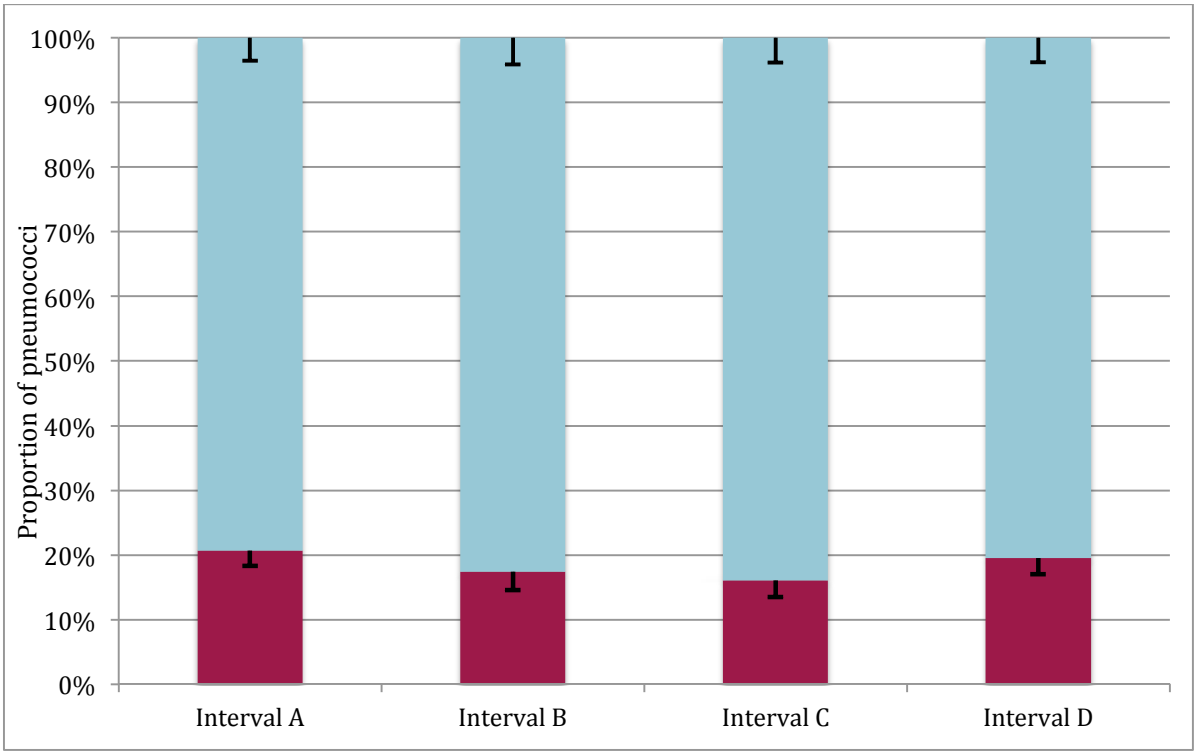

Figure 37: Proportion of vaccine (red bar graphs) and non-vaccine type (blue bar graphs) pneumococci. Interval A = denotes the period preceding 1st dose of PCV-13. Interval B = period after 1st but prior to the 2nd PCV-13 dose. Interval C = period after 2nd dose but before 3rd PCV-13 dose while interval D represents the period after the 3rd dose of PCV-13

5.4.6.2 Is early life carriage (before first dose of PCV-13) associated with increased risk of pneumococcal acquisition after the first but before the second dose of PCV-13?

Based on the descriptive summary above and the schema in Figure 39, a Cox proportional hazard model was applied to investigate the effect of prior pneumococcal colonisation on risk of acquiring any (either VT or NVT) pneumococci following each sequential dose of PCV-13. Children who were colonised with any type of pneumococcus prior to their 1st dose of PCV-13 (interval A) were 2.4 times (95% CI 1.5 - 3.9, $p < 0.001$) more likely to acquire any type of pneumococcus after the 1st dose (interval B) compared to those who were not colonised prior to the 1st dose.

When prior colonisation was stratified into either VT or NVT, children who were colonised by VT or NVT prior to the 1st dose (interval A) were 3.2 times (95% CI 1.4 - 7.3, $p = 0.004$) and 2.7 times (95% CI 1.6 - 4.6, $p < 0.001$) more likely to be colonised by any type of pneumococcus after the 1st dose (interval B) compared to those who were not colonised prior to their 1st dose respectively.

By applying a Fine & Gray competing risk model, it was shown that prior VT colonisation had no significant effect on the time to VT carriage acquisition after the 1st dose (interval B), HR = 1.97 (95% CI 0.25 – 15.10, $p = 0.51$). In contrast, prior NVT colonisation increased the hazard of NVT carriage acquisition after the 1st dose (interval B), HR = 3.57 (95% CI 1.95 – 6.52, $p < 0.001$), however this may simply reflect increased statistical power in the larger NVT group.

5.4.6.3 Is early life carriage (before second dose of PCV-13) associated with increased risk of pneumococcal acquisition after the second but before the third dose of PCV-13?

There was no increased risk of pneumococcal carriage acquisition in the interval between 2nd and 3rd dose (interval C) for children who were colonised before the 1st dose (interval A), HR = 1.1 (95% CI 0.8 – 1.76, $p = 0.38$). In contrast, children who were colonised in the interval between the 1st and 2nd dose (interval B) were 1.78 (95% CI 1.22– 2.57, $p = 0.003$) more likely to acquire any type of pneumococci after the 2nd dose of PCV-13 (interval C) compared to non-colonised children.

When prior colonisation episodes were stratified into either VT or NVT as above, children colonised by VT pneumococci either before the 1st dose (interval A) or between the 1st and 2nd dose (interval B) were not at a significant risk of acquiring any type of pneumococcus after the 2nd dose of PCV-13 compared to children who were not

colonised, HR = 1.99 (95% CI 0.94 – 4.18, p = 0.06) and HR = 1.59 (95% CI 0.95 – 2.65, p = 0.07) respectively. Similarly, children colonised with a NVT before the 1st dose (interval A) were not at a significant risk of acquiring any type of pneumococcus after the 2nd dose of PCV-13 compared to children who were not colonised, HR = 1.99 (95% CI 0.80 – 1.97, p = 0.32). In contrast, children colonised with NVT between the 1st and 2nd dose (interval B) were 1.60 (95% CI 1.10 – 2.39, p = 0.01) times more likely to acquire any type of pneumococcus post 2nd dose of PCV-13 (but before 3rd dose of PCV-13) compared to children who were not colonised.

By applying the Fine & gray model, it was shown that prior colonisation with either VT or NVT had no significant effect (p > 0.1 in all instances) on risk of VT or NVT colonisation between the 2nd and 3rd dose (interval C). Similarly, prior NVT carriage had no effect (p > 0.2 in all instances) on acquisition of NVT pneumococci between the 2nd and 3rd dose (interval C).

5.4.7 Risk factors associated with pneumococcal carriage acquisition

Due to the differences in the population profiles at TC Newman and Mbekweni, Table 17 above, crude and adjusted incidence and hazard ratios were computed using Cox proportional hazard models to compare the association of each of the risk factors with the incidence of pneumococcal carriage acquisition. The presence of other siblings in a household was positively associated with acquisition of any type of pneumococcus at any given time-point, irrespective of clinical site, HR = 1.21 (95% CI 1.03 – 1.41, p < 0.02) table 18. Additional risk factors independently associated with acquisition of any type of pneumococcal serotype were weather (autumn) (HR = 1.38, 95% CI 1.15-1.65, p

< 0.001); fathers and older siblings who slept in the same room as the child (HR = 1.19, 95% CI 1.01-1.40, p = 0.03 and HR = 1.31, 95% CI 1.12-1.54, p < 0.001 respectively).

Table 17: Risk factors associated with the incidence of acquisition of carriage of **ANY** pneumococcal serotype irrespective of site

Risk factors	Carriage Events	Person-time (days)	Incidence rate (95% CI; episodes per 100 child day)	Hazard ratio (95% CI*)	*p*
Site					
Mbekweni	528	27162	1.94 (1.78-2.11)	1	-
TC Newman	399	20987	1.90 (1.72-2.10)	0.87 (0.71-1.07)	0.18
Gender					
Female	517	27320	1.89 (1.73-2.06)	1	-
Male	410	20829	1.97 (1.78-2.17)	1.03 (0.88-1.19)	0.71
Mode of delivery					
Normal vaginal	743	37984	1.95 (1.82-2.10)	1	-
Vacuum	2	381	0.52 (0.13-2.10)	0.38 (0.09-1.59)	0.18
Elective caesarean	65	3565	1.82 (1.42-2.32)	1.00 (0.75-1.33)	1.00
Emergency caesarean	117	6219	1.88 (1.57-2.25)	0.94 (0.75-1.18)	0.61
HIV exposure					
No	704	36491	1.93 (1.79-2.08)	1	-
Yes	223	11658	1.91 (1.68-2.18)	0.91 (0.75-1.10)	0.32
Feed choice					
Exclusive breastfeeding	301	14900	2.02 (1.80-2.26)	1	-
Mixed feeding	163	7773	2.09 (1.79-2.44)	1.03 (0.81-1.32)	0.79
Never breastfeeding	463	25533	1.81 (1.66-1.98)	0.96 (0.80-1.16)	0.67
Other Children[1]					
No	563	31918	1.76 (1.62-1.92)	1	-
Yes	364	16231	2.24 (2.02-2.49)	**1.21 (1.03-1.41)**	**0.02**
Premature (Mbekweni)					
No	470	23338	2.01 (1.83-2.20)	1	-
Yes	58	3824	1.51 (1.17-1.96)	0.74 (0.54-1.01)	0.06
Premature (TC Newman)					
No	329	17858	1.84 (1.65-2.05)	1	-
Yes	70	3129	2.23 (1.77-2.82)	1.12 (0.53-2.39)	-[#]
Season					
Summer	213	12194	1.75 (1.53-1.99)	1	-
Spring	196	12024	1.63 (1.41-1.87)	0.92 (0.75-1.12)	0.41
Winter	232	12213	1.89 (1.67-2.16)	1.07 (0.88-1.30)	0.46
Autumn	286	11718	**2.44 (2.17-2.74)**	**1.38 (1.15-1.65)**	**0.00**
Smoking status of Mother					
No	358	19147	1.87 (1.69-2.07)	1	-
Yes	569	29002	1.96 (1.81-2.13)	1.12 (0.94-1.34)	0.19
Sleeps with Mother					
No	108	4907	2.20 (1.82-2.63)	1	-
Yes	819	43242	1.89 (1.77-2.03)	0.81 (0.55-1.19)	0.29
Sleeps with Father					
No	526	29208	1.89 (1.65-1.96)	1	-
Yes	401	18941	**2.11 (1.91-2.33)**	**1.19 (1.01-1.40)**	**0.03**
Sleeps with sibling					
No	493	29033	1.69 (1.55-1.85)	1	-
Yes	434	19116	**2.27 (2.06-2.49)**	**1.31 (1.12-1.54)**	**0.00**
Microwave					
No	180	8001	2.24 (1.94-2.63)	1	-
Yes	747	40148	1.86 (1.73-1.99)	0.68 (0.45-1.03)	0.07
Paraffin					
No	843	44280	1.90 (1.78-2.04)	1	-
Yes	84	3869	2.17 (1.75-2.69)	0.85 (0.54-1.34)	0.48
Wood					
No	903	47136	1.91 (1.79-2.04)	1	-
Yes	24	1013	2.37 (1.59-3.53)	0.99 (0.56-1.73)	0.96
Gas					
No	904	46382	1.95 (1.82-2.08)	1	-
Yes	23	1767	1.30 (0.86-1.96)	0.74 (0.46-1.22)	0.24

*One observation with excluded for missing data. [1]Other children staying with the child in the same household. "-" Insufficient data in that category to make any inferences. [#]The *p*-value couldn't be computed directly since the hazard ratio was found through an interaction effect with site.

A multivariate analysis of risk factors independently associated with VT acquisition showed that although the presence of older siblings increased the hazard of acquiring a VT pneumococci, HR = 1.29 (95% CI 0.73-2.27), this effect was not statistically significant, p = 0.38, table 19. The only risk factors independently associated with VT pneumococcal acquisition were fathers and older siblings who slept in the same room as the child (HR = 1.95, 95% CI 1.07-3.53, p = 0.03 and HR = 1.88, 95% CI 1.07-3.30, p = 0.03 respectively).

Table 18: Risk factors associated with the incidence of carriage of **PCV-13** vaccine serotype pneumococci irrespective of site

Risk factors	**Carriage Events**	**Person-time (days)**	**Incidence rate (95% CI; episodes per 100 child day)**	**Hazard ratio (95% CI*)**	***p***
Site					
Mbekweni	51	17209	0.29 (0.22-0.40)	1	-
TC Newman	47	14803	0.31 (0.23-0.42)	1.00 (0.48-2.06)	1.00
Gender					
Female	51	18240	0.28 (0.21-0.37)	1	-
Male	47	13772	0.34 (0.26-0.45)	1.66 (0.97-2.84)	0.07
Mode of delivery					
Normal vaginal	80	25859	0.31 (0.25-0.39)	1	-
Vacuum	1	382	0.26 (0.034-1.8)	2.56 (0.21-31.59)	0.47
Elective caesarean	9	2356	0.38 (0.19-0.73)	2.01 (0.80-5.04)	0.14
Emergency caesarean	8	3415	0.23 (0.12-0.47)	0.62 (0.23-1.73)	0.37
HIV exposure					
No	73	23587	0.31 (0.25-0.39)	1	-
Yes	25	8425	0.30 (0.20-0.44)	0.87 (0.44-1.70)	0.68
Feed choice					
Exclusive breastfed	38	10672	0.36 (0.26-0.49)	1	-
Mixed fed	17	4043	0.42 (0.26-0.68)	1.27 (0.54-3.01)	0.59
Never breastfed	43	17297	0.25 (0.18-0.34)	0.76 (0.40-1.47)	0.42
Other Children[1]					
No	58	21150	0.27 (0.21-0.35)	1	-
Yes	40	10862	0.37 (0.27-0.50)	1.29 (0.73-2.27)	0.38
Premature					
No	80	27406	0.29 (0.23-0.36)	1	-
Yes	18	4606	0.39 (0.24-0.62	1.04 (0.52-2.11)	0.91
Season					
Summer	27	9510	0.28 (0.19-0.41)	1	-
Spring	19	8681	0.22 (0.41-0.34)	0.71 (0.34-1.50)	0.38
Winter	22	6050	0.36 (0.24-0.55)	1.54 (0.773.11)	0.23
Autumn	30	7771	0.39 (0.27-0.55)	1.07 (0.56-2.04)	0.84
Smoking status of Mother					
No	39	13033	0.30 (0.21-0.41)		
Yes	59	18979	0.31 (0.24-0.40)	0.87 (0.45-1.71)	0.70
Sleeps with Mother					
No	11	3367	0.33 (0.18-0.59)	1	-

Yes	87	28645	0.30 (0.25-0.37)	4.60 (0.31-68.96)	0.27
Sleeps with Father					
No	49	20269	0.24 (0.18-0.32)	1	-
Yes	49	11743	**0.42 (0.32-0.55)**	**1.95 (1.07-3.53)**	**0.03**
Sleeps with sibling					
No	49	20258	0.24 (0.18-0.32)	1	-
Yes	49	11754	**0.41 (0.31-0.55)**	**1.88 (1.07-3.30)**	**0.03**
Microwave					
No	23	4905	0.47 (0.31-0.71)	1	-
Yes	75	27107	**0.28 (0.22-0.35)**	**0.08 (0.01-1.23)**	**0.07**
Paraffin					
No	89	29993	0.30 (0.24-0.37)	1	-
Yes	9	2019	0.45 (0.23-0.86)	0.21 (0.01-3.81)	0.29
Wood					
No	95	31750	0.30 (0.24-0.37)	1	-
Yes	3	262	1.14 (0.37-3.55)	1.79 (0.09-36.74)	0.71
Gas					
No	97	31171	0.31 (0.26-0.38)	1	-
Yes	1	841	0.12 (0.02-0.84)	0.08 (0.00-1.47)	0.09

*One observation with excluded for missing data. [1]Other children staying with the child in the same household. "-" Insufficient data in that category to make any inferences. [#]The p-value couldn't be computed directly since the hazard ratio was found through an interaction effect with site. Looking at the confidence interval, we can see that this hazard ratio is insignificant.

A multivariate analysis of risk factors independently associated with NVT acquisition showed significant interaction between feeding choice and clinic site. Therefore, the relationship between feeding choice and NVT carriage acquisition is presented by site. Compared to children from Mbekweni who were exclusively breastfed for the first 6 months of life, children who were never breastfed were less likely to acquire NVT pneumococci (HR = 0.60 [95% CI 0.43-0.83, $p < 0.001$]), table 20. In contrast, children from TC Newman who were both breastfed and formula fed (mixed fed) were more likely to acquire NVT pneumococci compared to children who were exclusively breast (HR = 1.68 [95% CI 1.08-2.62, $p = 0.02$]). Additional risk factors independently associated with acquisition of NVT pneumococci were weather (autumn) (HR = 1.44, 95% CI 1.16-1.80, $p < 0.001$); children staying with mothers with self-reported tobacco smoking (HR = 1.26, 95% CI 1.01-1.57, $p = 0.04$) and older siblings who slept in the same room as the child (HR = 1.31, 95% CI 1.07-1.60, $p = 0.01$), table 20.

Table 19: Risk factors associated with the incidence of carriage of **non-PCV-13 vaccine** serotypes irrespective of site

Risk factors	Carriage Events	Person-time (days)	Incidence rate (95% CI; episodes per 100 child day)	Hazard ratio (95% CI*)	*p*
Site					
Mbekweni	382	27222	1.40 (1.27-1.55)	1	-
TC Newman	280	20870	1.34 (1.19-1.51)	0.87 (0.68-1.12)	0.28
Gender					
Female	373	27336	1.36 (1.23-1.51)	1	-
Male	289	20756	1.39 (1.24-1.56)	1.00 (0.83-1.21)	0.98
Mode of delivery					
Normal vaginal	531	37825	1.40 (1.29-1.53)	1	-
Vacuum	1	382	0.26 (0.03-1.85)	0.27 (0.03-2.14)	0.22
Elective caesarean	43	3575	1.20 (0.89-1.62)	0.83 (0.57-1.21)	0.34
Emergency caesarean	87	6310	1.38 (1.11-1.70)	1.01 (0.76-1.33)	0.96
HIV exposure					
No	507	36492	1.39 (1.27-1.51)	1	-
Yes	155	11600	1.34 (1.14-1.56)	0.84 (0.65-1.08)	0.18
Feed choice (Mbekweni)					
Exclusive breastfed	107	6276	1.70 (1.41-2.06)	1	-
Mixed fed	76	5323	1.43 (1.14-1.79)	0.72 (0.49-1.05)	0.09
Never breastfed	199	15623	1.27 (1.10-1.46)	**0.60 (0.43-0.83)**	**0.00**
Feed choice (TC Newman)					
Exclusive breastfed	106	8321	1.27 (1.05-1.54)	1	-
Mixed fed	47	2461	1.91 (1.43-2.54)	**1.68 (1.08-2.62)**	**0.02**
Never breastfed	127	10088	1.26 (1.06-1.50)	1.01 (0.72-1.40)	0.97
Other Children[1]					
No	398	32063	1.24 (1.13-1.37)	1	-
Yes	264	16029	1.65 (1.46-1.86)	**1.19 (0.98-1.44)**	**0.09**
Premature					
No	570	41193	1.38 (1.27-1.50)	1	-
Yes	92	6899	1.33 (1.09-1.63)	0.96 (0.72-1.26)	0.75
Season					
Summer	155	13251	1.17 (1.00-1.37)	1	-
Spring	145	11728	1.24 (1.05-1.45)	1.03 (0.81-1.31)	0.80
Winter	172	12106	1.42 (1.22-1.65)	1.16 (0.93-1.46)	0.20
Autumn	190	11007	1.73 (1.50-1.99)	**1.44 (1.16-1.80)**	**0.00**
Smoking status of Mother					
No	250	19107	1.31 (1.16-1.48)	1	-
Yes	412	28985	1.42 (1.29-1.57)	**1.26 (1.01-1.57)**	**0.04**
Sleeps with Mother					
No	81	5003	1.62 (1.30-2.01)	1	-
Yes	581	43089	1.35 91.24-1.46)	0.73 (0.45-1.18)	0.19
Sleeps with Father					
No	388	28998	1.33 (1.21-1.48)	1	-
Yes	274	19094	**1.44 (1.27-1.61)**	1.05 (0.85-1.29)	0.65
Sleeps with sibling					
No	358	29085	1.23 (1.11-1.37)	1	-
Yes	304	19007	**1.60 (1.43-1.79)**	**1.31 (1.07-1.60)**	**0.01**
Microwave					
No	128	8105	1.58 (1.32-1.88)	1	-
Yes	534	39987	**1.34 (1.22-1.45)**	0.77 (0.46-1.29)	0.32
Paraffin					
No	603	44216	1.36 (1.26-1.48)	1	-
Yes	59	3876	1.52 (1.18-1.96)	1.01 (0.57-1.80)	0.97
Wood					
No	644	47077	1.37 (1.27-1.48)	1	-
Yes	18	1015	1.77 (1.12-2.81)	1.19 (0.59-2.40)	0.63
Gas					
No	649	46322	1.40 (1.29-1.51)	1	-
Yes	13	1770	0.73 (0.42-1.26)	0.54 (0.27-1.04)	0.07

5.5 DISCUSSION

In this intensively sampled birth cohort study, we show a gradual increase in the incidence of pneumococcal carriage acquisition from 1.2 episodes per 100 child-days to 1.8 episodes per 100 child-days and finally 2.0 episodes per 100 child-days at the time of the first, second and third PCV-13 dose respectively. Pneumococcal colonisation either prior to the first dose or after the first but before the second dose of PCV-13 was associated with an increased hazard for subsequent pneumococcal colonisation, HR = 2.4 (95% CI 1.50-3.90, $p < 0.001$) and HR = 1.78 (95% CI 1.22-2.57, $p = 0.003$) respectively. This finding is consistent with data from other studies that showed that prior VT colonisation was associated with reduced PCV efficacy against serotypes included in the vaccine formulations (60,61,81). This was characterized by serotype hypo-responsiveness to PCV-7 in infants carrying pneumococcus prior to or at the time of immunisation. Dagan *et al.* showed that VT pneumococcal carriage amongst Israeli children at the time of, or shortly before, PCV-7 immunisation was associated with serotype hypo-responsiveness to the carried VT pneumococcal capsular polysaccharide (61). Similarly, Madhi *et al.* demonstrated that VT carriage amongst South African children at any stage prior to completion of sequential PCV-7 immunisation schedule was associated with inferior qualitative and quantitative serotype-specific antibody responses to the colonising strain of pneumococcus (81). Our data however needs further validation with a larger sample size. In addition, such carriage data will need to be complimented by data from serotype-specific functional antibody titers.

The exact mechanism behind serotype hyporesponsiveness remains incompletely understood. It has however been suggested that early life carriage in healthy infants

results in high capsular polysaccharide loads that may induce B-cell exhaustion or unresponsiveness (61,66,82). Serotype-specific immune response to the pneumococcal polysaccharide is T-cell independent and incapable of generating functional B-cell immunologic memory (37). Therefore, early life pneumococcal carriage may lead to immature B-cells that do not effectively differentiate into antibody producing plasma cells in response to PCV immunisation (83,84). The turn-over of insufficiently differentiated plasma B-cells can also deplete the serotype-specific B-cell pool within the marginal zone of the spleen and nodes (83,84).

PCV immunisation reduces NP colonisation by VT serotypes, and increase NVT colonisation within months following receipt of primary vaccine series (11,39,85–88). Further, the degree to which PCV confers protection against VT carriage acquisition depends on pneumococcal carriage prevalence and the immunisation coverage within communities. In our study, 'residual' carriage of VT pneumococci did not appear to be affected by sequential doses of PCV-13. This contrasts earlier randomized controlled trials and observational studies (most of which were conducted prior to community-wide introduction of PCVs) from different setting and using different vaccine formulations that showed reduced VT colonisation (11,39,85–88). A randomized, double-blinded placebo controlled trial conducted by Mbelle *et al* using PCV-9 administered at 6, 10 and 14 weeks in 500 infants from Soweto showed significant reductions in overall VT NP carriage (18 vs 36%, $p < 0.001$) in vaccinees at age 9 months and was accompanied by an increase in NVT NP carriage (36 vs 25%, $p < 0.001$) (11). There are limited data on the long-term effect of each sequential PCV dose on the likelihood of VT acquisition post community-wide implementation of PCVs in routine immunisation programs. Lakshman *et al* found no reduction in VT NP colonisation a few

years after primary series of PCV-7 and PPV-23 booster immunisation (89). Similar findings were shown in a study conducted in our settings by Madhi *et al* who followed-up 352 South African children who participated in a PCV-9 randomized controlled trial described above. In this study, PCV-9 no longer reduced VT or NVT NP colonisation in children 5.3 years after receipt of third dose of PCV-9 during infancy (90). The residual VT carriage in our setting and the lack of evidence to suggest reductions in VT colonisation may be a function of (1) high community-wide PCV-13 vaccine coverage resulting in significant changes in the dynamics of VT community-wide exposure (herd protection) and risk of colonisation (2) incompletely immunised older siblings or adults with waning immunity (particularly amongst HIV-infected adults) may serve as reservoirs for increased risk of person-to-person transmission. The lack of catch-up immunisation in older children and low immunogenicity of pneumococcal serotypes presents constant challenges in maintaining highly protective antibody titres in early life. Our study was however not sufficiently powered to explore the impact of PCV on individual pneumococcal serotypes as well as determine the potential reservoir for the circulating residual VT pneumococci.

We did not observe any differences in susceptibility to VT pneumococci between HIV-exposed uninfected (HEU) and HIV-unexposed uninfected (HUU) children. This corroborates earlier findings from a longitudinal (CIPRA 04) study amongst children from Soweto (Gauteng) and Tygerberg (Western cape) which showed that PCV-7 had similar effect against VT colonisation in HEU, HUU and HIV infected children during their first 18 months of life (91). Some studies In contrast, Gill *et al* suggested that HEU children are at increased risk pneumococcal colonisation, especially serogroups 6, 19 and 23, than HUU (92,93). The differences in susceptibility to pneumococcal

colonisation may be attributed to inconsistencies in implementation of HIV management practices especially presumptive co-trimoxazole prophylaxis.

We found no association between breastfeeding and pneumococcal carriage acquisition in general. However, children who were never breastfed from Mbekweni were less likely to acquire NVT pneumococci than children who were exclusively breastfed (HR=0.60 [95% CI 0.43-0.83], $p < 0.001$). The association between breast-feeding and pneumococcal carriage in human and in vitro studies is largely inconclusive (94,95). Some studies have shown that breastfeeding protects against respiratory tract infections as well as IPD. Lower rates of pneumococcal carriage have been shown in exclusively breastfed children while another study suggests no association between breastfeeding and pneumococcal carriage. There are several factors that may explain these differences (1) behavior – HIV infected mothers may less likely breastfeed their children in order to minimize the risk of postpartum HIV transmission to infants. This corroborated our findings as well as others that suggests that HEU infants from Mbekweni, a setting with a high HIV prevalence, are less likely to be colonised by NVT compared to HUU (HR = 0.60 [95% CI 0.43-0.83, $p < 0.001$]) (2) other, unmeasured demographic factors

Several risk factors including daycare center attendance, viral respiratory tract infections, living in a crowded family (especially the presence of older siblings), lower socioeconomic status, use of antibiotics and exposure to tobacco smoke predispose to nasopharyngeal carriage of pneumococci (67–70,72,77). In our settings however, the presence of other children that was associated with VT and NVT carriage acquisition irrespective of site. In addition, environmental tobacco smoke exposure was associated

with NVT carriage acquisition and this effect was more pronounced at TC Newman where 50% of mothers reported tobacco smoking compared to 9% observed amongst mothers at Mbekweni. Evidence suggests that tobacco smoke exposure decreases mucocilliary clearance and induces the inflammation of the respiratory epithelial lining thereby predisposing individuals to bacterial adherence and subsequent colonisation (96,97). Similar finding have been reported amongst children staying in a crowded refugee camp in Thailand (26). Our finding that tobacco smoking is associated with NVT acquisition is different from what was reported by Greenberg *et al.* who showed that tobacco smoking as rather associated with VT carriage and not NVT (98). This study was however conducted in PCV naïve Israeli children. Prior to PCV implementation, children were more likely to be colonised by VT pneumococci whereas in the post PCV era, children are more frequently colonised by NVT as shown in our cohort. This is also consistent with the observed high proportion of NVT observed in our cohort.

Study limitations include the inability to identify co-colonisation with multiple serotypes and the lack of MLST data. Therefore, we cannot speculate whether all the NVT detected in the present study result from expansion of pre-existing clones (unmasking), introduction of new clones not previously encountered in our setting or acquisition of novel capsular types through partial or complete exchange of the capsular biosynthetic gene cluster (capsular switching). Such a process would generally result in persistence of VT pneumococci in the form of mutants expressing NVT capsular types. Another limitation of the present study is the fact that tobacco exposure was based on self-reported smoking which might be inaccurate. Inclusion of biochemical measures of passive smoking such as urinary cotinine levels might provide a more objective measure of the amount of tobacco exposure in children (99). However, another study

has shown that urinary cotinine level did not improve the validity of tobacco smoke exposure and its relation to lower respiratory illness in infants questionnaires (100).

In summary, early life NP colonisation was associated with an increased hazard for subsequent pneumococcal colonisation post PCV-13 immunisation. The season (autumn), presence of older siblings in the household and fathers who slept in the same room as the child were associated with VT pneumococcal carriage acquisition irrespective of site. On the other hand, infant mode of feeding within the first six years of life, season (autumn), self-reported maternal smoking and older children in the household who actually slept in the same room as the infant were associated with NVT pneumococcal carriage acquisition.

5.6 REFERENCES

1. Cutts FT, Zaman SMA, Enwere G, Jaffar S, Levine OS, Okoko JB, 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) [Internet]. Jan [cited 2015 Sep 30];365(9465):1139–46.
2. Prymula R, Peeters P, Chrobok V, Kriz P, Novakova E, Kaliskova E, et al. Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both *Streptococcus pneumoniae* and non-typable Haemophilus influenzae: a randomised double-blind efficacy study. Lancet [Internet]. 2006 Mar 4 [cited 2015 Sep 29];367(9512):740–8.
3. Eskola J, Kilpi T, Palmu A, Jokinen J, Haapakoski J, Herva E, et al. Efficacy of a pneumococcal conjugate vaccine against acute otitis media. N Engl J Med [Internet]. 2001 Mar 8 [cited 2015 Sep 29];344(6):403–9.
4. Klugman KP, Madhi SA, Huebner RE, Kohberger R, Mbelle N, Pierce N. A trial of a 9-valent pneumococcal conjugate vaccine in children with and those without HIV infection. N Engl J Med [Internet]. 2003 Oct 2 [cited 2015 Sep 29];349(14):1341–8.
5. O'Brien KL, Moulton LH, Reid R, Weatherholtz R, Oski J, Brown L, et al. Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial. Lancet [Internet]. 2003 Aug 2 [cited 2015 Sep 29];362(9381):355–61.
6. Ruiz-Aragón J, Márquez Peláez S, Molina-Linde JM, Grande-Tejada AM. Safety and immunogenicity of 13-valent pneumococcal conjugate vaccine in infants: a meta-analysis. Vaccine [Internet]. 2013 Nov 4 [cited 2015 Aug 14];31(46):5349–58.
7. Black S, Shinefield H, Fireman B, Lewis E, Ray P, Hansen JR, 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 [Internet]. 2000 Mar [cited 2015 Sep 16];19(3):187–95.
8. Steens A, Caugant DA, Aaberge IS, Vestrheim DF. Decreased Carriage and Genetic Shifts in the *Streptococcus pneumoniae* Population After Changing the 7-Valent to the 13-Valent Pneumococcal Vaccine in Norway. Pediatr Infect Dis J [Internet].

2015 May 26 [cited 2015 Aug 10];

9. Obaro SK, Adegbola RA, Banya WA, Greenwood BM. Carriage of pneumococci after pneumococcal vaccination. Lancet (London, England) [Internet]. 1996 Jul 27 [cited 2015 Sep 30];348(9022):271–2.
10. Dagan R, Givon-Lavi N, Zamir O, Sikuler-Cohen M, Guy L, Janco J, et al. Reduction of nasopharyngeal carriage of *Streptococcus pneumoniae* after administration of a 9-valent pneumococcal conjugate vaccine to toddlers attending day care centers. J Infect Dis [Internet]. 2002 Apr 1 [cited 2015 Sep 22];185(7):927–36.
11. Mbelle N, Huebner RE, Wasas AD, Kimura A, Chang I, Klugman KP. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. J Infect Dis [Internet]. 1999 Oct [cited 2015 Sep 30];180(4):1171–6.
12. Dagan R, Givon-Lavi N, Fraser D, Lipsitch M, Siber GR, Kohberger R. Serum serotype-specific pneumococcal anticapsular immunoglobulin g concentrations after immunization with a 9-valent conjugate pneumococcal vaccine correlate with nasopharyngeal acquisition of pneumococcus. J Infect Dis [Internet]. 2005 Aug 1 [cited 2015 Sep 22];192(3):367–76.
13. O'Brien KL, Millar E V, Zell ER, Bronsdon M, Weatherholtz R, Reid R, et al. Effect of pneumococcal conjugate vaccine on nasopharyngeal colonization among immunized and unimmunized children in a community-randomized trial. J Infect Dis [Internet]. 2007 Oct 15 [cited 2015 Feb 2];196(8):1211–20.
14. Simell B, Auranen K, Käyhty H, Goldblatt D, Dagan R, O'Brien KL. The fundamental link between pneumococcal carriage and disease. Expert Rev Vaccines [Internet]. 2012 Jul [cited 2015 Feb 20];11(7):841–55.
15. Moore MR, Hyde TB, Hennessy TW, Parks DJ, Reasonover AL, Harker-Jones M, et al. Impact of a Conjugate Vaccine on Community-Wide Carriage of Nonsusceptible *Streptococcus pneumoniae* in Alaska. J Infect Dis [Internet]. 2004 Dec 1 [cited 2015 Sep 30];190(11):2031–8.
16. Millar E V, O'Brien KL, Watt JP, Bronsdon MA, Dallas J, Whitney CG, et al. Effect of community-wide conjugate pneumococcal vaccine use in infancy on nasopharyngeal carriage through 3 years of age: a cross-sectional study in a high-risk population. Clin Infect Dis [Internet]. 2006 Jul 1 [cited 2015 Sep 30];43(1):8–15.

17. Huang SS, Hinrichsen VL, Stevenson AE, Rifas-Shiman SL, Kleinman K, Pelton SI, et al. Continued impact of pneumococcal conjugate vaccine on carriage in young children. Pediatrics [Internet]. 2009 Jul [cited 2015 Aug 31];124(1):e1–11.
18. Rudan I, O'Brien KL, Nair H, Liu L, Theodoratou E, Qazi S, et al. Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. J Glob Health [Internet]. 2013 Jun [cited 2014 May 1];3(1):010401.
19. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet [Internet]. 2012 Jun 9 [cited 2014 Jul 10];379(9832):2151–61.
20. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet. Elsevier Ltd; 2010 Jun;375(9730):1969–87.
21. Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, et al. Global burden of childhood pneumonia and diarrhoea. Lancet [Internet]. 2013 Apr 20 [cited 2014 Aug 2];381(9875):1405–16.
22. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet [Internet]. 2009 Sep 12 [cited 2014 Oct 22];374(9693):893–902.
23. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet [Internet]. 2011 Dec 3 [cited 2014 May 11];378(9807):1962–73.
24. Hill PC, Cheung YB, Akisanya A, Sankareh K, Lahai G, Greenwood BM, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian infants: a longitudinal study. Clin Infect Dis [Internet]. 2008 Mar 15 [cited 2015 Sep 11];46(6):807–14.
25. Abdullahi O, Karani A, Tigoi CC, Mugo D, Kungu S, Wanjiru E, et al. Rates of acquisition and clearance of pneumococcal serotypes in the nasopharynges of children in Kilifi District, Kenya. J Infect Dis [Internet]. 2012 Oct 1 [cited 2015 Mar 4];206(7):1020–9.
26. Turner P, Turner C, Jankhot A, Helen N, Lee SJ, Day NP, et al. A longitudinal study

of *Streptococcus pneumoniae* carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. PLoS One [Internet]. 2012 Jan [cited 2015 Mar 4];7(5):e38271.

27. Wright A, Parry Morgan W, Colebrook L, Dodgson RW. Observations ON Prophylactic Inoculation Against Pneumococcus Infections, And On The Results Which Have Been Achieved By It. Lancet [Internet]. Elsevier; 1914 Jul 7 [cited 2015 Oct 22];183(4714):1–10.
28. Klugman PKP, Hayden Smith SW, Koornhof HJ. Evidence that prevention of carriage by pneumococcal capsular vaccines may be the mechanism of protection from pneumococcal pneumonia [Internet]. Southern African Journal of Infectious Diseases. 2011 [cited 2015 Oct 3]. p. 221–4.
29. Austrian R. A brief history of pneumococcal vaccines. Drugs Aging [Internet]. 1999 Jan [cited 2015 Feb 18];15 Suppl 1:1–10.
30. Austrian R. Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. Rev Infect Dis [Internet]. Jan [cited 2015 Jul 15];3 Suppl:S1–17.
31. Watson DA, Musher DM, Jacobson JW, Verhoef J. A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. Clin Infect Dis [Internet]. 1993 Nov [cited 2015 Jul 15];17(5):913–24.
32. Tomczyk S, Bennett NM, Stoecker C, Gierke R, Moore MR, Whitney CG, et al. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged ≥65 years: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep [Internet]. 2014 Sep 19 [cited 2015 Nov 4];63(37):822–5.
33. Weintraub A. Immunology of bacterial polysaccharide antigens. Carbohydr Res [Internet]. 2003 Nov 14 [cited 2015 Aug 12];338(23):2539–47.
34. Temple K, Greenwood B, Inskip H, Hall A, Koskela M, Leinonen M. Antibody response to pneumococcal capsular polysaccharide vaccine in African children. Pediatr Infect Dis J [Internet]. 1991 May [cited 2015 Nov 20];10(5):386–90.
35. Douglas RM, Hansman D, Miles HB PJ. Pneumococcal carriage and type-specific antibody. Am J Dis Child. 1986;(140):1183–5.
36. Koskela M, Leinonen M, Häivä VM, Timonen M, Mäkelä PH. First and second dose antibody responses to pneumococcal polysaccharide vaccine in infants. Pediatr

Infect Dis [Internet]. Jan [cited 2015 Nov 20];5(1):45–50.

37. Jakobsen H, Hannesdottir S, Bjarnarson SP, Schulz D, Trannoy E, Siegrist C-A, et al. Early life T cell responses to pneumococcal conjugates increase with age and determine the polysaccharide-specific antibody response and protective efficacy. Eur J Immunol [Internet]. 2006 Feb [cited 2015 Oct 17];36(2):287–95.
38. Giebink GS. The prevention of pneumococcal disease in children. N Engl J Med [Internet]. 2001 Oct 18 [cited 2015 Jun 25];345(16):1177–83.
39. Dagan R, Givon-Lavi N, Zamir O, Sikuler-Cohen M, Guy L, Janco J, et al. Reduction of nasopharyngeal carriage of *Streptococcus pneumoniae* after administration of a 9-valent pneumococcal conjugate vaccine to toddlers attending day care centers. J Infect Dis [Internet]. 2002 Apr 1 [cited 2015 Sep 22];185(7):927–36.
40. Borgoño JM, McLean AA, Vella PP, Woodhour AF, Canepa I, Davidson WL, et al. Vaccination and revaccination with polyvalent pneumococcal polysaccharide vaccines in adults and infants. Proc Soc Exp Biol Med [Internet]. 1978 Jan [cited 2015 Oct 2];157(1):148–54.
41. Sell SH, Wright PF, Vaughn WK, Thompson J, Schiffman G. Clinical studies of pneumococcal vaccines in infants. I. Reactogenicity and immunogenicity of two polyvalent polysaccharide vaccines. Rev Infect Dis [Internet]. Jan [cited 2015 Nov 20];3 Suppl:S97–107.
42. Avci FY, Li X, Tsuji M, Kasper DL. A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. Nat Med [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011 Dec [cited 2015 Nov 20];17(12):1602–9.
43. Pletz MW, Maus U, Krug N, Welte T, Lode H. Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. Int J Antimicrob Agents [Internet]. 2008 Sep [cited 2015 Nov 21];32(3):199–206.
44. Avci FY. Novel strategies for development of next-generation glycoconjugate vaccines. Curr Top Med Chem [Internet]. 2013 Jan [cited 2015 Feb 18];13(20):2535–40.
45. Goldblatt D, Hussain M, Andrews N, Ashton L, Virta C, Melegaro A, et al. Antibody responses to nasopharyngeal carriage of *Streptococcus pneumoniae* in adults: a longitudinal household study. J Infect Dis [Internet]. 2005 Aug 1 [cited 2015 Aug 31];192(3):387–93.

46. Simell B, Korkeila M, Pursiainen H, Kilpi TM, Käyhty H. Pneumococcal carriage and otitis media induce salivary antibodies to pneumococcal surface adhesin a, pneumolysin, and pneumococcal surface protein a in children. J Infect Dis [Internet]. 2001 Mar 15 [cited 2015 Sep 24];183(6):887–96.
47. Weinberger DM, Dagan R, Givon-Lavi N, Regev-Yochay G, Malley R, Lipsitch M. Epidemiologic evidence for serotype-specific acquired immunity to pneumococcal carriage. J Infect Dis [Internet]. 2008 Jun 1 [cited 2015 Sep 24];197(11):1511–8.
48. Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis [Internet]. 2010 Jan 1 [cited 2015 Apr 23];201(1):32–41.
49. Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis [Internet]. 2011 Oct [cited 2015 Aug 11];11(10):760–8.
50. Melegaro A, Choi YH, George R, Edmunds WJ, Miller E, Gay NJ. Dynamic models of pneumococcal carriage and the impact of the Heptavalent Pneumococcal Conjugate Vaccine on invasive pneumococcal disease. BMC Infect Dis. 2010 Jan;10:90.
51. Rudan I. Epidemiology and etiology of childhood pneumonia. Bull World Health Organ. 2008 May;86(5):408–16.
52. Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. Lancet Infectious Diseases. Elsevier; 2005. p. 83–93.
53. Richards L, Ferreira DM, Miyaji EN, Andrew PW, Kadioglu A. The immunising effect of pneumococcal nasopharyngeal colonisation; protection against future colonisation and fatal invasive disease. Immunobiology [Internet]. 2010 Apr [cited 2015 Nov 20];215(4):251–63.
54. McCool TL, Cate TR, Moy G, Weiser JN. The immune response to pneumococcal proteins during experimental human carriage. J Exp Med [Internet]. 2002 Feb 4 [cited 2015 Sep 22];195(3):359–65.
55. Wright AKA, Ferreira DM, Gritzfeld JF, Wright AD, Armitage K, Jambo KC, et al. Human nasal challenge with *Streptococcus pneumoniae* is immunising in the absence of carriage. PLoS Pathog [Internet]. Public Library of Science; 2012 Jan 5 [cited 2015 Nov 20];8(4):e1002622.

56. Snapper CM, Shen Y, Khan AQ, Colino J, Zelazowski P, Mond JJ, et al. Distinct types of T-cell help for the induction of a humoral immune response to *Streptococcus pneumoniae.* Trends Immunol [Internet]. 2001 Jun [cited 2015 Nov 20];22(6):308–11.
57. Levine O, Greenwood B. Opportunities and challenges for pneumococcal conjugate vaccines in low-and middle-income countries. … vaccines impact … [Internet]. 2008 [cited 2015 Sep 28];
58. O'Brien KL, Hochman M, Goldblatt D. Combined schedules of pneumococcal conjugate and polysaccharide vaccines: is hyporesponsiveness an issue? Lancet Infect Dis [Internet]. 2007 Sep [cited 2015 Sep 28];7(9):597–606.
59. Borrow R, Stanford E, Waight P, Helbert M, Balmer P, Warrington R, et al. Serotype-specific immune unresponsiveness to pneumococcal conjugate vaccine following invasive pneumococcal disease. Infect Immun [Internet]. 2008 Nov [cited 2015 Sep 28];76(11):5305–9.
60. Väkeväinen M, Soininen A, Lucero M, Nohynek H, Auranen K, Mäkelä PH, et al. Serotype-specific hyporesponsiveness to pneumococcal conjugate vaccine in infants carrying pneumococcus at the time of vaccination. J Pediatr [Internet]. 2010 Nov [cited 2015 Sep 28];157(5):778–83.e1.
61. Dagan R, Givon-Lavi N, Greenberg D, Fritzell B, Siegrist C-A. Nasopharyngeal carriage of *Streptococcus pneumoniae* shortly before vaccination with a pneumococcal conjugate vaccine causes serotype-specific hyporesponsiveness in early infancy. J Infect Dis [Internet]. 2010 May 15 [cited 2015 Sep 28];201(10):1570–9.
62. Madhi SA, Violari A, Klugman KP, Lin G, McIntyre JA, von Gottberg A, et al. Inferior quantitative and qualitative immune responses to pneumococcal conjugate vaccine in infants with nasopharyngeal colonization by *Streptococcus pneumoniae* during the primary series of immunization. Vaccine [Internet]. 2011 Sep [cited 2015 Sep 28];29(40):6994–7001.
63. Givon-Lavi N, Dagan R, Fraser D, Yagupsky P, Porat N. Marked differences in pneumococcal carriage and resistance patterns between day care centers located within a small area. Clin Infect Dis [Internet]. 1999 Nov [cited 2015 Sep 24];29(5):1274–80.
64. Barnes DM, Whittier S, Gilligan PH, Soares S, Tomasz A, Henderson FW.

Transmission of multidrug-resistant serotype 23F *Streptococcus pneumoniae* in group day care: evidence suggesting capsular transformation of the resistant strain in vivo. J Infect Dis [Internet]. 1995 Apr [cited 2016 Jan 8];171(4):890–6.

65. Weinberger DM, Grant LR, Steiner CA, Weatherholtz R, Santosham M, Viboud C, et al. Seasonal drivers of pneumococcal disease incidence: impact of bacterial carriage and viral activity. Clin Infect Dis [Internet]. 2014 Jan [cited 2015 Sep 24];58(2):188–94.
66. Gray BM, Turner ME, Dillon HC. Epidemiologic studies of *Streptococcus pneumoniae* in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. Am J Epidemiol [Internet]. 1982 Oct [cited 2015 Feb 20];116(4):692–703.
67. Abdullahi O, Nyiro J, Lewa P, Slack M, Scott JAG. The descriptive epidemiology of *Streptococcus pneumoniae* and *Haemophilus influenzae* nasopharyngeal carriage in children and adults in Kilifi district, Kenya. Pediatr Infect Dis J [Internet]. 2008 Jan [cited 2015 Sep 24];27(1):59–64.
68. Syrjänen RK, Kilpi TM, Kaijalainen TH, Herva EE, Takala a K. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Finnish children younger than 2 years old. J Infect Dis. 2001 Aug;184(4):451–9.
69. Sleeman KL, Griffiths D, Shackley F, Diggle L, Gupta S, Maiden MC, et al. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. J Infect Dis [Internet]. 2006 Sep 1 [cited 2015 Aug 27];194(5):682–8.
70. Sjöström K, Spindler C, Ortqvist A, Kalin M, Sandgren A, Kühlmann-Berenzon S, et al. Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. Clin Infect Dis [Internet]. 2006 Feb 15 [cited 2015 Sep 24];42(4):451–9.
71. Bogaert D, De Groot R, Hermans PWM. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis [Internet]. 2004 Mar [cited 2014 May 1];4(3):144–54.
72. Weinberger DM, Simonsen L, Jordan R, Steiner C, Miller M, Viboud C. Impact of the 2009 influenza pandemic on pneumococcal pneumonia hospitalizations in the United States. J Infect Dis [Internet]. 2012 Feb 1 [cited 2015 Sep 24];205(3):458–65.

73. Walaza S, Tempia S, Dawood H, Variava E, Moyes J, Cohen AL, et al. Influenza virus infection is associated with increased risk of death amongst patients hospitalized with confirmed pulmonary tuberculosis in South Africa, 2010¿2011. BMC Infect Dis [Internet]. 2015 Jan 27 [cited 2015 Feb 28];15(1):26.
74. Wolter N, Tempia S, Cohen C, Madhi SA, Venter M, Moyes J, et al. High nasopharyngeal pneumococcal density, increased by viral coinfection, is associated with invasive pneumococcal pneumonia. J Infect Dis [Internet]. 2014 Nov 15 [cited 2014 Dec 15];210(10):1649–57.
75. O'Brien KL, Walters MI, Sellman J, Quinlisk P, Regnery H, Schwartz B, et al. Severe pneumococcal pneumonia in previously healthy children: the role of preceding influenza infection. Clin Infect Dis [Internet]. 2000 May [cited 2015 Sep 24];30(5):784–9.
76. Oddy WH, Sly PD, de Klerk NH, Landau LI, Kendall GE, Holt PG, et al. Breast feeding and respiratory morbidity in infancy: a birth cohort study. Arch Dis Child [Internet]. 2003 Mar [cited 2015 Sep 24];88(3):224–8.
77. Lynch JP, Zhanel GG. *Streptococcus pneumoniae*: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. Curr Opin Pulm Med [Internet]. 2010 May [cited 2015 Jan 25];16(3):217–25.
78. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2014.
79. Hastie T, Tibshirani R, Friedman J. Elements of Statistical Learning: Data Mining, Inference and Prediction. Second Edi. New York: Springer-Verlag; 2009. 745 p.
80. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. Ann Stat [Internet]. Institute of Mathematical Statistics; 1982 Dec 1 [cited 2015 Sep 23];10(4):1100–20.
81. Madhi SA, Violari A, Klugman KP, Lin G, McIntyre JA, von Gottberg A, et al. Inferior quantitative and qualitative immune responses to pneumococcal conjugate vaccine in infants with nasopharyngeal colonization by *Streptococcus pneumoniae* during the primary series of immunization. Vaccine [Internet]. 2011 Sep 16 [cited 2015 Sep 28];29(40):6994–7001.
82. Ekdahl K, Ahlinder I, Hansson HB, Melander E, Mölstad S, Söderström M, et al. Duration of nasopharyngeal carriage of penicillin-resistant *Streptococcus pneumoniae:* experiences from the South Swedish Pneumococcal Intervention

Project. Clin Infect Dis [Internet]. 1997 Nov [cited 2015 Oct 17];25(5):1113–7.

83. Timens W, Boes A, Rozeboom-Uiterwijk T, Poppema S. Immaturity of the human splenic marginal zone in infancy. Possible contribution to the deficient infant immune response. J Immunol [Internet]. 1989 Nov 15 [cited 2015 Oct 5];143(10):3200–6.
84. Siegrist C-A, Aspinall R. B-cell responses to vaccination at the extremes of age. Nat Rev Immunol [Internet]. 2009 Mar [cited 2015 Jul 4];9(3):185–94.
85. Cheung Y-B, Zaman SMA, Nsekpong ED, Van Beneden CA, Adegbola RA, Greenwood B, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian children who participated in a 9-valent pneumococcal conjugate vaccine trial and in their younger siblings. Pediatr Infect Dis J [Internet]. 2009 Nov [cited 2015 Oct 19];28(11):990–5.
86. Millar E V, O'Brien KL, Watt JP, Bronsdon MA, Dallas J, Whitney CG, et al. Effect of community-wide conjugate pneumococcal vaccine use in infancy on nasopharyngeal carriage through 3 years of age: a cross-sectional study in a high-risk population. Clin Infect Dis [Internet]. 2006 Jul 1 [cited 2015 Sep 30];43(1):8–15.
87. Dagan R, Patterson S, Juergens C, Greenberg D, Givon-Lavi N, Porat N, 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 [Internet]. 2013 Oct [cited 2014 May 23];57(7):952–62.
88. Prymula R, Kriz P, Kaliskova E, Pascal T, Poolman J, Schuerman L. Effect of vaccination with pneumococcal capsular polysaccharides conjugated to Haemophilus influenzae-derived protein D on nasopharyngeal carriage of *Streptococcus pneumoniae* and H. influenzae in children under 2 years of age. Vaccine [Internet]. 2009 Dec 10 [cited 2015 Oct 14];28(1):71–8.
89. Lakshman R, Murdoch C, Race G, Burkinshaw R, Shaw L, Finn A. Pneumococcal nasopharyngeal carriage in children following heptavalent pneumococcal conjugate vaccination in infancy. Arch Dis Child [Internet]. 2003 Mar [cited 2016 Jan 21];88(3):211–4.
90. Madhi SA, Adrian P, Kuwanda L, Cutland C, Albrich WC, Klugman KP. Long-term effect of pneumococcal conjugate vaccine on nasopharyngeal colonization by

Streptococcus pneumoniae--and associated interactions with Staphylococcus aureus and Haemophilus influenzae colonization--in HIV-Infected and HIV-uninfected children. J Infect Dis [Internet]. 2007 Dec 1 [cited 2016 Jan 21];196(11):1662–6.

91. Madhi SA, Izu A, Nunes MC, Violari A, Cotton MF, Jean-Philippe P, et al. Longitudinal study on *Streptococcus pneumoniae*, Haemophilus influenzae and Staphylococcus aureus nasopharyngeal colonization in HIV-infected and -uninfected infants vaccinated with pneumococcal conjugate vaccine. Vaccine [Internet]. 2015 Apr 21 [cited 2015 Apr 27];
92. Gill CJ, Mwanakasale V, Fox MP, Chilengi R, Tembo M, Nsofwa M, et al. Effect of presumptive co-trimoxazole prophylaxis on pneumococcal colonization rates, seroepidemiology and antibiotic resistance in Zambian infants: a longitudinal cohort study. Bull World Health Organ [Internet]. 2008 Dec [cited 2015 Dec 23];86(12):929–38.
93. Gill CJ, Mwanakasale V, Fox MP, Chilengi R, Tembo M, Nsofwa M, et al. Impact of human immunodeficiency virus infection on *Streptococcus pneumoniae* colonization and seroepidemiology among Zambian women. J Infect Dis [Internet]. 2008 Apr 1 [cited 2015 Sep 11];197(7):1000–5.
94. Andersson B, Porras O, Hanson LA, Lagergård T, Svanborg-Edén C. Inhibition of attachment of *Streptococcus pneumoniae* and *Haemophilus influenzae* by human milk and receptor oligosaccharides. J Infect Dis [Internet]. 1986 Feb [cited 2016 Jan 10];153(2):232–7.
95. Rosen IA, Håkansson A, Aniansson G, Hansson C, Andersson B, Nylén O, et al. Antibodies to pneumococcal polysaccharides in human milk: lack of relationship to colonization and acute otitis media. Pediatr Infect Dis J [Internet]. 1996 Jun [cited 2016 Jan 10];15(6):498–507.
96. Jones JG, Minty BD, Lawler P, Hulands G, Crawley JC, Veall N. Increased alveolar epithelial permeability in cigarette smokers. Lancet (London, England) [Internet]. 1980 Jan 12 [cited 2015 Oct 13];1(8159):66–8.
97. Piatti G, Gazzola T, Allegra L. Bacterial adherence in smokers and non-smokers. Pharmacol Res [Internet]. 1997 Dec [cited 2015 Oct 13];36(6):481–4.
98. Greenberg D, Givon-Lavi N, Broides A, Blancovich I, Peled N, Dagan R. The Contribution of Smoking and Exposure to Tobacco Smoke to *Streptococcus*

pneumoniae and *Haemophilus influenzae* Carriage in Children and Their Mothers. Clin Infect Dis [Internet]. 2006 Apr 1 [cited 2015 Sep 22];42(7):897–903.

99. Bakoula CG, Kafritsa YJ, Kavadias GD, Lazopoulou DD, Theodoridou MC, Maravelias KP, et al. Objective passive-smoking indicators and respiratory morbidity in young children. Lancet (London, England) [Internet]. 1995 Jul 29 [cited 2015 Oct 13];346(8970):280–1.

100. Margolis PA, Keyes LL, Greenberg RA, Bauman KE, LaVange LM. Urinary cotinine and parent history (questionnaire) as indicators of passive smoking and predictors of lower respiratory illness in infants. Pediatr Pulmonol [Internet]. 1997 Jun [cited 2015 Oct 13];23(6):417–23.

CHAPTER 6

Respiratory pathogens in children hospitalised with suspected pulmonary tuberculosis in Cape Town, South Africa

6.1 INTRODUCTION

Lower respiratory tract infection is a leading cause of mortality and morbidity in children under five years accounting for approximately 1 million deaths in 2013, with the majority in children younger than five years (1). Pulmonary tuberculosis (PTB) is an important cause of lower respiratory tract infections and may present as either an acute or chronic infection (2). TB is increasingly recognized as a primary cause or underlying comorbidity amongst children with pneumonia, particularly in settings endemic for human immunodeficiency virus (HIV) and tuberculosis (3).

TB diagnosis in children is largely made on the basis of clinical and radiological presentations, which may be non-specific. Therefore TB cannot be easily differentiated from other causes of acute or chronic respiratory tract infection (4). Recent studies have reported an increase in the detection of polymicrobial infections in children with respiratory tract infections, including children with pulmonary TB (5-7). Culture confirmed TB was reported in 8% of children hospitalised with acute pneumonia in South Africa, with no difference by HIV status; similar findings have been reported in other African studies(8). A recent meta-analysis confirmed *Mycobacterium tuberculosis* in 8% of childhood pneumonia cases in high TB endemic areas (3).

The availability of a real-time, fluorescent probe-based reverse transcription polymerase chain reaction (RT-PCR), has led to an exponential increase in our understanding of the diversity of polymicrobial respiratory infections which often have indistinguishable clinical presentations (6,9-20). It is the method of choice for sensitive detection and precise quantification of minute amounts of nucleic acid from potential respiratory tract pathogens. In addition, RT-PCR offers the ability to amplify and differentiate multiple targets within a single reaction tube. A number of both

laboratory-developed and commercial multiplex RT-PCR assays employing different amplification platforms have been described (21-29). These assays offer a high diagnostic yield, detect fastidious or non-culturable microbes within clinically relevant turn-around times, and have the potential to provide supplementary information, such as the presence of antibiotic resistance genes (18,21,22,25,26,28,30-33). Multiplex RT-PCR assays are nonetheless costly, require specialised laboratory equipment and use highly multiplexed reactions that may be deficient in individual assay performance and can be difficult to modify without extensive assay re-optimisation. The Fast-track Diagnostics (FTD) respiratory pathogens multiplex assay kit (FTD Resp33 kit, Fast-track Diagnostics, Luxembourg) utilises a standard commercial one-step RT-PCR amplification reaction with differentially labelled target-specific hydrolysis probes. The assay can easily be performed on most common real-time PCR instruments. A single assay is comprised of eight multiplex RT-PCR reactions for the detection of 33 potential respiratory pathogens. Similar FTD assays have been adopted for the study of polymicrobial infections elsewhere although mostly targeting fewer than 21 potential respiratory pathogens (23,29,34,35).

Little is known about the range and role of respiratory pathogens that may be detected in children presenting with suspected TB, either as the primary cause of respiratory symptoms in children who do not have TB, or as co-pathogens in children who do have TB. To date, only one study in adults has addressed this issue (36). Lockman *et al*., utilising microscopy and culture, showed that among TB suspects in Botswana, TB was confirmed in 52%; 17% had acute *Mycoplasma pneumoniae* infection, and 3% had *Pneumocystis jirovecii* infection. Co-infection with two or more pathogens was documented in 25% of patients. No other respiratory pathogen was found in 48% of patients with negative Acid Fast Bacillus (AFB) microscopy (36).

In our cohort studies of South African children hospitalised with suspected pulmonary TB (PTB), 15% had microbiologically-confirmed PTB, 48% were classified as having possible PTB and in 37% PTB was excluded, with respiratory disease due to other pathogens (37). Approximately half of children were treated for PTB, including all children with definite PTB and most with possible PTB (37). There is therefore a large amount of unaccounted respiratory illness amongst this cohort, as well as possible co-infection with *M. tuberculosis* (MTB) and other pathogens. We therefore conducted a study to describe the diversity of potential respiratory pathogens in children hospitalised with suspected PTB and to explore co-infection patterns between MTB and other respiratory pathogens.

6.2 Aims and Objectives

We aim to describe the diversity of potential respiratory pathogens in children hospitalised with suspected PTB and to explore co-infection patterns between MTB and other respiratory pathogens. Specific objectives are:

1. To identify the diversity of potential respiratory pathogens in children hospitalised with suspected pulmonary TB in whom:
    a. TB was microbiologically confirmed
    b. TB was excluded
2. To assess the age-dependent microbial pathogen distribution among children.
3. To explore co-infection patterns between MTB and other respiratory pathogens, and clustering of pathogens using quadratic discriminant analysis methodology.

6.3 MATERIALS AND METHODS

6.3.1 Study population

Children under 15 years of age suspected of PTB presenting at Red Cross War Memorial Children's Hospital (RCH) in Cape Town, between July 2011 and May 2012 were considered for enrolled in the TB diagnostic study, as previously described (10). RCH is a specialist referral paediatric hospital that manages approximately 250 000 patient visits a year. Verbal assent was obtained from children older than 7 years of age and informed consent was obtained from a parent or legal guardian for children of all ages for HIV testing and enrollment. Children with HIV infection were classified according to WHO clinical staging from stage 1 to stage 4 (38).

Suspected TB was defined as having history of a cough of 14 days or more plus one of the following: i) a household contact infected with tuberculosis within the previous three months, ii) loss of weight or failure to gain weight in the previous three months, iii) a positive tuberculin skin test (TST) to purified protein derivative (PPD; 2TU, PPD RT23, Staten Serum Institute, Denmark, Copenhagen), or iv) a chest radiograph suggestive of pulmonary tuberculosis. A positive skin test was defined as 5 mm or more of transverse induration in children with HIV infection or 10 mm or more in children without HIV infection (39). Children were excluded if the child was on TB treatment or TB prophylaxis for more than 72 h, and patients for whom follow-up was judged to be difficult (e.g. patients residing outside Cape Town and patients who could not produce adequate induced sputum). All patients with laboratory confirmed TB and those strongly suspected of having TB based on clinical and radiological criteria were referred for TB therapy at a local TB clinic in accordance with South African National Guidelines (40). Children were followed up at 1 and 3 months to assess response to therapy or

improvement without TB treatment. All children received care in the public health system that includes free EPI immunisation for DPT, HiB, PCV13 etc – you need to provide basic EPI schedule for SA as pathogens may depend on immunisation provided. (PCV7 from 2009, replaced with PCV13 in 2011 at 6, 14 weeks and 9 months.

6.3.2 Collection of respiratory specimens

Children with signs and symptoms suggestive of TB were categorised as definite TB (i.e. culture confirmed), not TB (i.e. not clinically diagnosed with improvement on follow-up without TB treatment) and others (i.e. possible TB). Two paired induced sputum and nasopharyngeal (NP) swab specimens were collected and transported to the laboratory within 2 h of collection. NP swabs were obtained before sputum induction using nylon flocked swabs (Copan Italia, Brescia, Italy) by trained study staff (41).

6.3.3 Processing of induced sputum for mycobacterial culture and confirmation

Induced sputum (IS) was collected by a trained research nurse at least 30 m after obtaining an NP swab. This was timed after subjecting the study participants to a 2 – 3 h fast in a well-ventilated room, as previously described (42). The respiratory specimens were then submitted to the National Health Laboratory Services (NHLS) Microbiology Laboratory at Groote Schuur Hospital (Cape Town, South Africa) for mycobacterial liquid culture (BACTEC MGIT, Becton Dickinson Microbiology Systems, Cockeysville, MD) and nucleic acid amplification testing (Xpert MTB/RIF, Cepheid, Sunnyvale, CA).

6.3.4 Processing of NP swabs for detection of respiratory pathogens

Each NP swab was immediately placed into 1.5 ml PrimeStore® transport and stabilization medium (PrimeStore® MTM, Longhorn Vaccines and Diagnostics, San Antonio, TX) and stored at –80°C until further batch processing as shown below (figure 40). We randomly selected samples from a convenience subset of 214 children for testing of NP specimens for other pathogens.

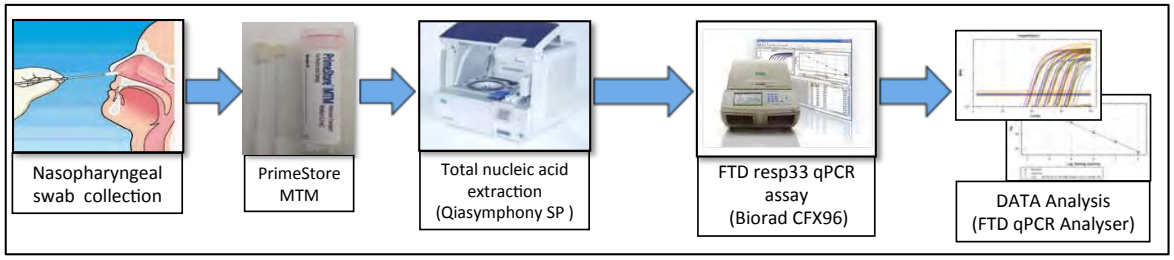

Figure 38: Specimen processing.

6.3.5 Total nucleic acid extraction

NP swabs stored in PrimeStore® MTM were thawed at room temperature (22°C) and vortexed for 15 s. Thereafter, 400 µl of each sample was transferred to a ZR BashingBeads™ Lysis tube (Zymo Research Corp., Irvine, CA) and subjected to a standardised "off board" mechanical lysis step on a Tissuelyzer LT (Qiagen, Hilden, Germany) (43). The lysed samples were then centrifuged at 10 000 g for 1 m to pellet all cellular debris. The supernatant was transferred to a sterile 2 ml reaction tube (Sarstedt, Nümbrecht, Germany) and 4 µl of an exogenous internal control (Equine arteritis virus) was added to each sample prior to automated total nucleic acid extraction on the QIAsymphony SP instrument using the QIAsymphony® Virus/Bacteria

mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions (REF). Total nucleic acid was eluted with 60 µl elution buffer and stored at –20°C until processing (figure 40).

6.3.6 Real-time multiplex PCR assay

The FTD Resp33 includes eight multiplex RT-PCR reactions per specimen for the detection of the following respiratory pathogens; Reaction 1: influenza virus A, influenza virus B and rhinovirus; Reaction 2: parainfluenza virus 2-4 and the equine arteritis virus (EAV) derived internal control; Reaction 3: coronaviruses 229E, NL63, O43 and HKU1; Reaction 4: parainfluenza virus 1, human metapneumovirus A/B, bocavirus, and *Mycoplasma pneumoniae*; Reaction 5: respiratory syncytial virus A/B, cytomegalovirus, enterovirus/parechovirus, and adenovirus (AV); Reaction 6: *Staphylococcus aureus*, *Chlamydophila pneumoniae*, *Haemophilus influenzae* type b and *Streptococcus pneumoniae*; Reaction 7: *Pneumocystis jirovecii*, *Legionella* species (Legio), *Klebsiella pneumoniae* and *Salmonella* species; Reaction 8: *Moraxella catarrhalis*, influenza C, *Bordetella pertussis* and *Haemophilus influenzae species*. The qPCR targets are shown in table 21.

Table 20: Target pathogens in the FTD respiratory pathogens 33 multiplex real-time PCR assay.

Multiplex Reaction No.	**Primer/ Probe mix#**	**Target pathogen (Nucleic acid type)**	**Target sequence genes**
1	FluAB-RH	Influenza A (**RNA**)	Matrix gene (*pos*1)
		Influenza B (**RNA**)	Segment 8 NS1/NEP
		Rhinovirus (**RNA**)	5'untranslated region (5' UTR)
2	Para-EAV	Parainfluenza 3 (**RNA**)	Hemagglutinin-neuraminidase (HN) mRNA
		Parainfluenza 2 (**RNA**)	Hemagglutinin-neuraminidase (HN) mRNA
		Parainfluenza 4 (**RNA**)	Hemagglutinin-neuraminidase (HN) mRNA
		Internal control*	

3	Cor	Coronavirus 229 (**RNA**)	Nucleocapsid protein (N) gene
		Coronavirus 63 (**RNA**)	Nucleocapsid protein (N) gene
		Coronavirus HKU1 (**RNA**)	Nucleocapsid protein (N) gene
		Coronavirus 43 (**RNA**)	Nucleocapsid protein (N) gene
4	BoMpPf1	Parainfluenza 1 (**RNA**)	Hemagglutinin-neuraminidase (HN) mRNA
		Human Metapneumoviruses A/B (**RNA**)	Fusion glycoprotein (F) gene
		Bocavirus (**DNA**)	Nonstructural protein (NP1) gene
		Mycoplasma pneumoniae (**DNA**)	Adhesin P1
5	RsEPAcmv	Respiratory syncytial viruses A/B (**RNA**)	Nucleocapsid protein (N) gene
		Cytomegalovirus (**DNA**)	US7+US8 genes
		Enterovirus (**RNA**)	Parts of domain IV and V
		Parechovirus (**RNA**)	5'untranslated region (UTR)
		Adenovirus (**DNA**)	Hexon gene
6	Bac	*Staphylococcus aureus* (**DNA**)	Sensor histidine kinase (*vicK*) gene
		Chlamydophila pneumoniae (**DNA**)	RNA polymerase beta chain gene
		Haemophilus influenza B (**DNA**)	capsular export (*BexA*) gene
		Streptococcus pneumoniae (**DNA**)	*LytA gene*
7	KLePSa	*Pneumocystis jirovecii* (**DNA**)	*(mtLSU) rRNA gene*
		Legionella spp (**DNA**)	16 S rRNA
		Klebsiella pneumoniae (**DNA**)	Hemolysin (*Khe*) gene
		Salmonella spp (**DNA**)	*Tetrathionate subunit B (ttrB)*
8	MoBoCH	*Moraxella catarrhalis* (**DNA**)	Outer membrane proteins (*cop*B) gene
		Influenza C (RNA)	Matrix gene
		Bordetella pertussis (**DNA**)	Insertion sequence (IS481)
		Haemophilus influenza spp (**DNA**)	Outer membrane proteins *(ompP2)* gene

Note. *Equine arteritis virus (EAV) based Internal control. #Pathogen names as indicated on package inserts for each multiplex reaction mix.

Briefly, the eight multiplex RT-PCR reactions were set-up on two 96-well PCR plates, with the first five multiplex reactions on plate 1 and the three remaining multiplex reactions on plate 2. Each multiplex reaction comprised 194 µl of 2× RT-PCR buffer with 23.3 µl of each primer/probe mix and 15.5 µl of 25× enzyme mix (AgPath-IDTM One-Step RT-PCR Kit, Applied Biosystems, Carlsbad, CA). Subsequently, 15 µl of each reaction mix was added to 14 wells of a PCR plate that included 12 sample reactions plus one positive and one negative control. The PCR reaction mix contained 10 µl of total nucleic acid extract or positive controls that was added to 15 µl of the respective reaction mix to a total volume of 25 µl. PCR amplification was performed on the Bio-Rad CFX96 Touch™ RT-PCR amplification system (Bio-Rad Laboratories, Hercules, CA). The thermal cycling conditions consisted of an initial denaturation at 95°C for 10 m,

followed by 40 amplification cycles of 95°C for 8 s (denaturation) and 60°C for 34 s (amplification and detection). Specimens were considered positive if they had a well-defined amplification curve that crossed the cycle threshold (Cq-) within 40 cycles. Assay performance was monitored by the positive, negative and internal amplification controls that were provided with the kit.

6.3.7 Development of FTD Resp33 qPCR Analyser

To minimize transcription errors when analysing the multiplex real-time RT-PCR results, a JAVA based program, FTD Resp 33 qPCR Analyser (figure 41), was developed and is freely available on the web (http://www.gematics.com/analyser.html). Further details on how to run the program are provided (Appendix 6.3 and snapshot of the excel output [figure S6.3]). The FTD Resp33 qPCR Analyser program automatically analyses FTD resp33 results from the Bio-Rad CFX96 Touch™ RT-PCR amplification system to circumvent transcription and transposition errors associated with rearranging large datasets into a table, or estimating the absolute quantities from external quantification standards for each of the 33 pathogens.

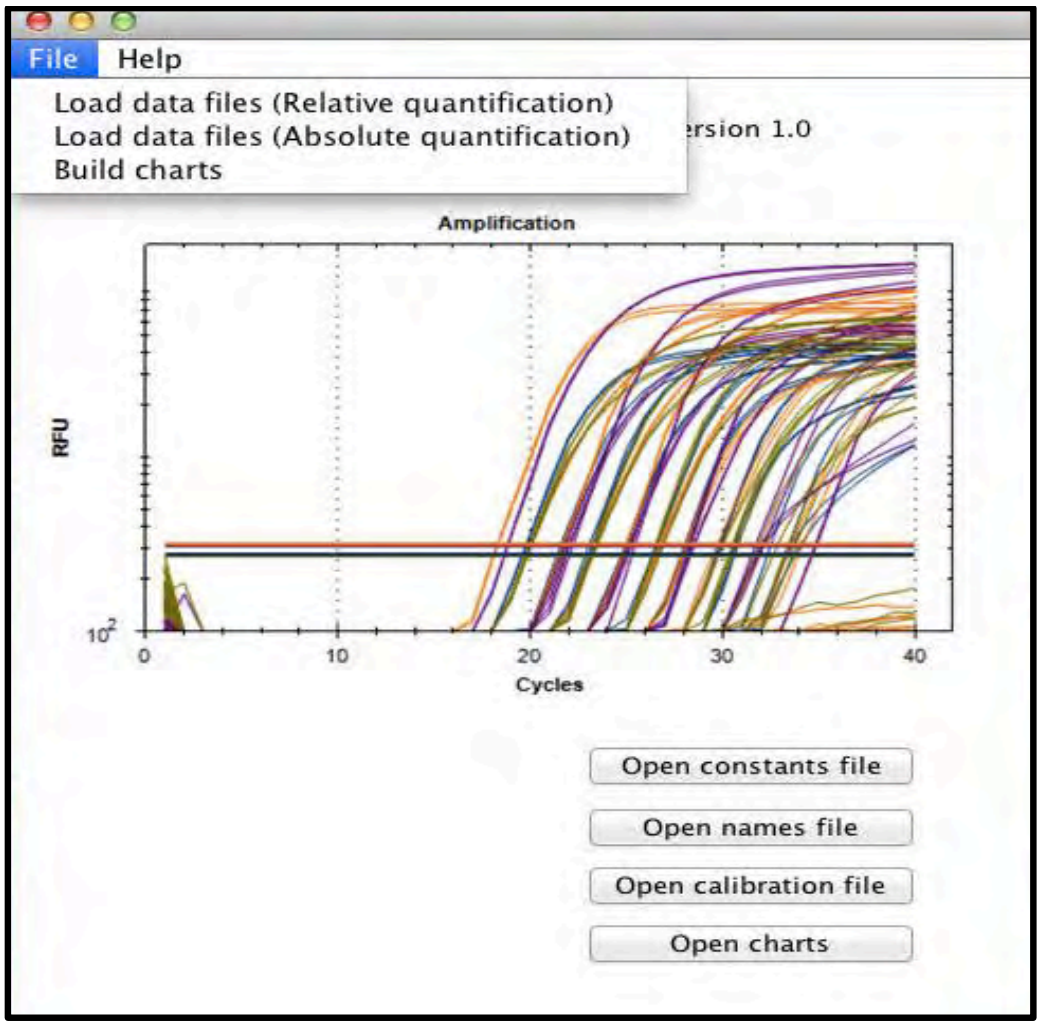

Figure 39: Screenshot of the FTD Resp33 qPCR Analyser program.

6.3.8 Statistical Considerations

Initial exploratory statistics were performed using STATA software (Stata Corporation, College Station, TX) and the rest of the analyses were performed using the openly available statistical environment R, version 3.1.1 (44). A logistic regression model, with age group as a predictor of the occurrence of each respiratory pathogen, was used to determine the age-dependent respiratory pathogen distribution. We used analysis of variance (ANOVA) to compare the occurrence of each pathogen across the three TB categories and the resulting significance (*p*-values) were adjusted for false discovery due to multiple comparisons by using the Benjamini & Hochberg method (45).

Permutation tests were used to determine which pathogen pairs were statistically concurrent. Briefly, for each pair of pathogens, *X* and *Y,* the observed number of concurrences (*m*) was counted. The null hypothesis was that there was no relationship between pathogens *X* and *Y* and that the co-occurrences were purely random. The null hypothesis was tested by generating random permutations of the occurrences of *X* and *Y*. The number of concurrences under these random conditions was then counted, m_1. By repeating the permutation process 10 000 times, 10 000 random concurrences were obtained as follows: m_1, m_2,…, m_{10000}. The shape of the null-distribution, the distribution of co-occurrence counts under purely random conditions, was estimated from the 10 000 observed values of the permutation test. A histogram for each pathogen pair was then constructed to determine the shape of the distribution. The achieved significance level (ASL) was computed as the number of permuted co-occurrences that were equal to or greater than the observed number of concurrence (tail probability under the null-distribution), this can be interpreted as a parametric *p*-value (46).

Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) (47) were used to optimally discriminate respiratory pathogens occurring in children with and without microbiologically confirmed TB, and those in whom TB was excluded. A Canonical Variate Analysis (CVA) biplot was used to provide a graphical representation of the LDA and a QDA biplot analysis (47). The QDA plot was constructed of *C. pneumoniae*, *M. pneumoniae*, RSV, Influenza C virus, coronavirus 043, *S. pneumoniae*, *P. jirovecii*, *H. influenzae* type b, *M. cartarrhalis*, *S. aureus*, Influenza B virus, enterovirus, *H. influenzae spp*, rhinovirus, cytomegalovirus and hMPV in order to obtain a visual representation of the observed differences in the definite TB, non-TB, and possible TB groups. For all the tests, a *p*-value less than 0.05 was used as the limit of statistical significance.

6.3.9 Ethical considerations

Written informed consent was obtained from a parent and/or legal guardian. This study was approved by the Human Research Ethics Committee, Faculty of Health Science, University of Cape Town (**HREC ref: 045/2008**) and the Provincial Government of the Western Cape, South Africa.

6.4 RESULTS

6.4.1 Patient characteristics and TB culture results.

Baseline characteristics of the 214 children enrolled are summarised in Table 22. The median age was 36 months (interquartile range, [IQR] 19 - 66 months), which did not significantly differ by TB category (p = 0.84), although children with definite TB tended to be older. Overall, 16% (34/214) of the children had culture confirmed TB, while 40% (86/214) had possible TB based on clinical and radiological findings and TB was excluded in 44% (94/214) of the children. Only 27 (13%) of children were HIV-infected, with similar prevalence by TB category. There was no difference in symptoms, including the median duration of cough across the three TB categories. Of the children in whom the immunisation status data was available, 129 (60%) had their national EPI upto-date whereas this information was missing in 36 (17%) and not reported in 47 (23%) children.

Table 21: Baseline characteristics of the children in the study by TB category

Characteristics	Total (n=214)	Definite TB (n=34)	Possible TB (n=86)	Not-TB (n=94)	*p*-value
Age in months, median (IQR)	36 (19 – 66)	56 (23 – 109)	34 (15 – 63)	32 (17 – 63)	0.003
Duration of cough in days, median (IQR)	15 (6 – 25)	20 (6 – 22)	17 (9 – 26)	14 (6 – 22)	0.36
Sex: female, n (%)	109 (51)	14 (41)	45 (52)	50 (53)	0.22
HIV infection, n (%)	27 (13)	7 (21)	13 (15)	7 (7)	0.14
HIV WHO clinical staging, n (%)*					
Stage 1	0	0	0	0	-
Stage 2	0	0	0	0	-
Stage 3	21 (10)	5 (15)	10 (12)	6 (6)	0.36
Stage 4	6 (3)	2 (6)	3 (35)	1 (1)	0.48

**Refers to the WHO HIV classification which is based on the clinical manifestation of disease in the absence of CD4 cell count measurements or other diagnostic and laboratory testing methods in resource constrained settings to inform patient care.*

6.4.2 Diversity of potential respiratory pathogens

Among the 214 nasopharyngeal samples, 97% (207/214) positive for at least one potential respiratory pathogen. *M. catarrhalis* (64%), *S. pneumoniae* (42%), *H. influenzae spp* (29%) and *S. aureus* (22%) were the most common bacterial pathogens detected (Table 23 and figure 42). *M. pneumoniae* (9%), *B. pertussis* (7%) or *C. pneumoniae* (4%) were detected less commonly. The most common viruses detected were hMPV (19%), rhinovirus (15%), influenza C (9%), adenovirus (7%), cytomegalovirus (7%) or coronavirus O43 (5.6%) (Table 23 and figure 43).

Table 22: Respiratory pathogens detected from a cohort of children presenting with suspected pulmonary tuberculosis[£]

	Total n (%)	**TB status n (%)**			**[#]Adjusted *p*-value**
		Definite TB	**Possible TB**	**Not-TB**	
Number of Participants	214 (100)	34 (16)	86 (40)	94 (44)	
Bacteria					
M. catarrhalis	137 (64)	24 (71)	56 (65)	57 (60)	
S. pneumoniae	90 (42)	18 (53)	38 (44)	34 (36)	
H. influenzae spp	62 (29)	8 (24)	25 (29)	29 (31)	
S. aureus	47 (22)	7 (21)	18 (21)	22 (23)	
H. influenzae B	31 (15)	5 (15)	15 (17)	11 (12)	
M. pneumoniae	19 (9)	2 (6)	9 (10)	8 (9)	
B. pertussis	12 (6)	3 (9)	6 (7)	3 (3)	
C. pneumoniae	9 (4)	4 (12)	4 (5)	1 (1)	0.42
Viruses					
Metapneumovirus A/B	41 (19)	7 (21)	21 (24)	13 (14)	
Rhinovirus	31 (15)	6 (18)	13 (15)	12 (13)	
Enterovirus	22 (10)	1 (3)	10 (12)	11 (12)	0.65
Adenovirus	14 (7)	3 (9)	5 (6)	6 (6)	
Cytomegalovirus	14 (7)	3 (9)	3 (3)	8 (9)	
Bocavirus	14 (7)	8 (24)	1 (1)	5 (5)	
Coronavirus	17 (8)	5 (15)	8 (9)	4 (4)	0.42
Coronavirus O43	12 (6)	5 (15)	5 (6)	2 (2)	0.42
Coronavirus HKU	2 (1)	0 (0)	1 (1)	1 (1)	
Coronavirus NL63	2 (1)	0 (0)	1 (1)	1 (1)	
Coronavirus 229E	1 (0.5)	0 (0)	1 (1)	0 (0)	
RSV A/B*	7 (3)	1 (3)	3 (3)	3 (3)	
Influenza virus	33 (15)	8 (24)	15 (17)	10 (11)	
Influenza A	7 (3)	1 (3)	4 (5)	2 (2)	
Influenza B	7 (3)	1 (3)	2 (2)	4 (4)	
Influenza C	19 (9)	6 (18)	9 (10)	4 (4)	
Parainfluenza virus	10 (5)	1 (3)	4 (5)	5 (5)	
Parainfluenza 1	5 (2)	1 (3)	2 (2)	2 (2)	
Parainfluenza 2	1 (0.5)	0 (0)	0 (0)	1 (1)	
Parainfluenza 3	2 (1)	0 (0)	1 (1)	1 (1)	
Parainfluenza 4	2 (1)	0 (0)	1 (1)	1 (1)	
Fungi					
P. jirovecii	23 (11)	3 (9)	10 (12)	10 (11)	

[£]*Arranged in order of decreasing colonisation or detection rates. *RSV A/B = Respiratory Syncytial virus A and B.* [#]The adjusted *p*-value compensates for false discovery multiple comparisons.

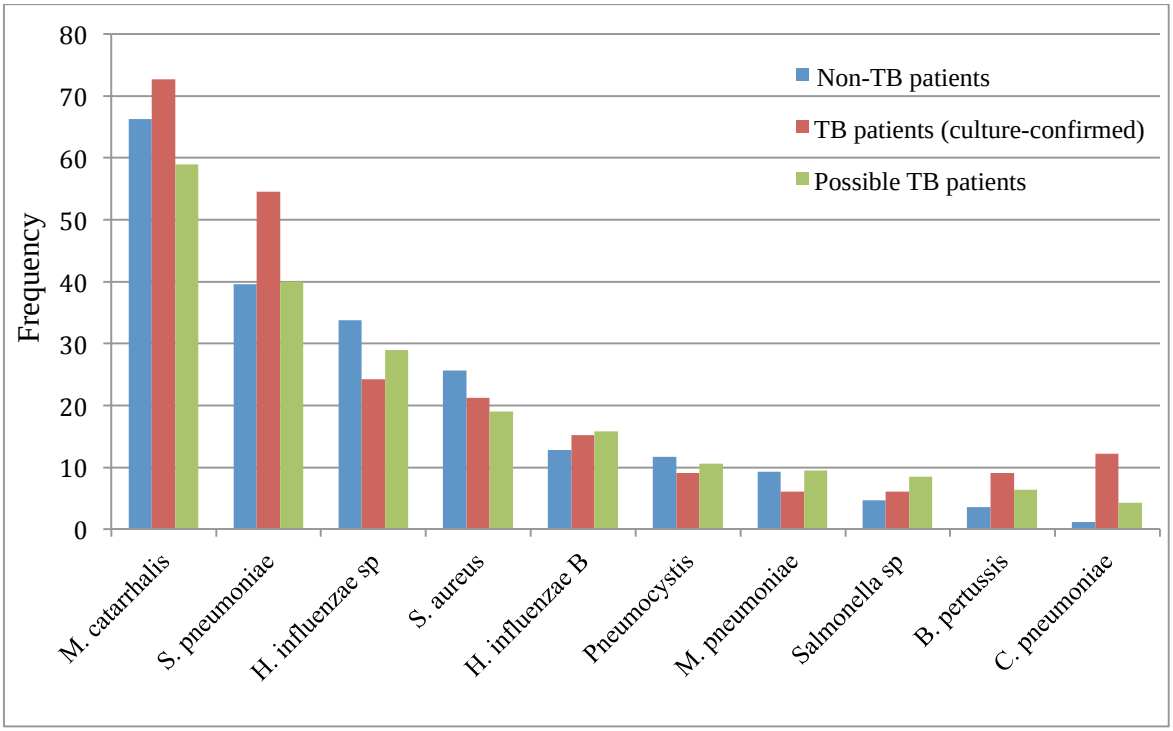

Figure 40: Diversity of bacterial and fungal pathogens stratified by TB category.

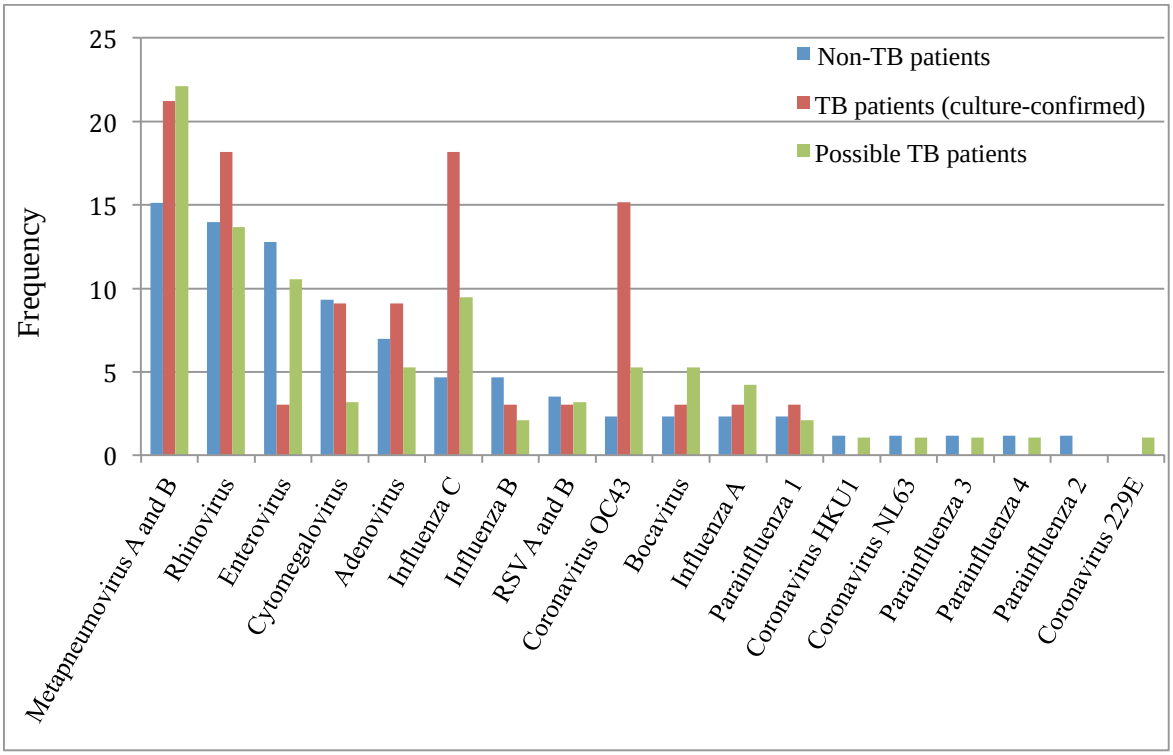

Figure 41: Diversity of viral respiratory pathogens stratified by TB category.

6.4.3 Age-dependent respiratory pathogen distribution

B. pertussis infection was mostly detected in children between 12 and 23 months old ($p < 0.01$) while bocavirus was mostly detected in children younger than three months ($p = 0.04$). (Figures 44 and 45, respectively). *M. catarrhalis,* rhinovirus, hMPV, influenza C virus, cytomegalovirus enterovirus and bocavirus were most commonly detected in children younger than three months (figures 44 and 45).

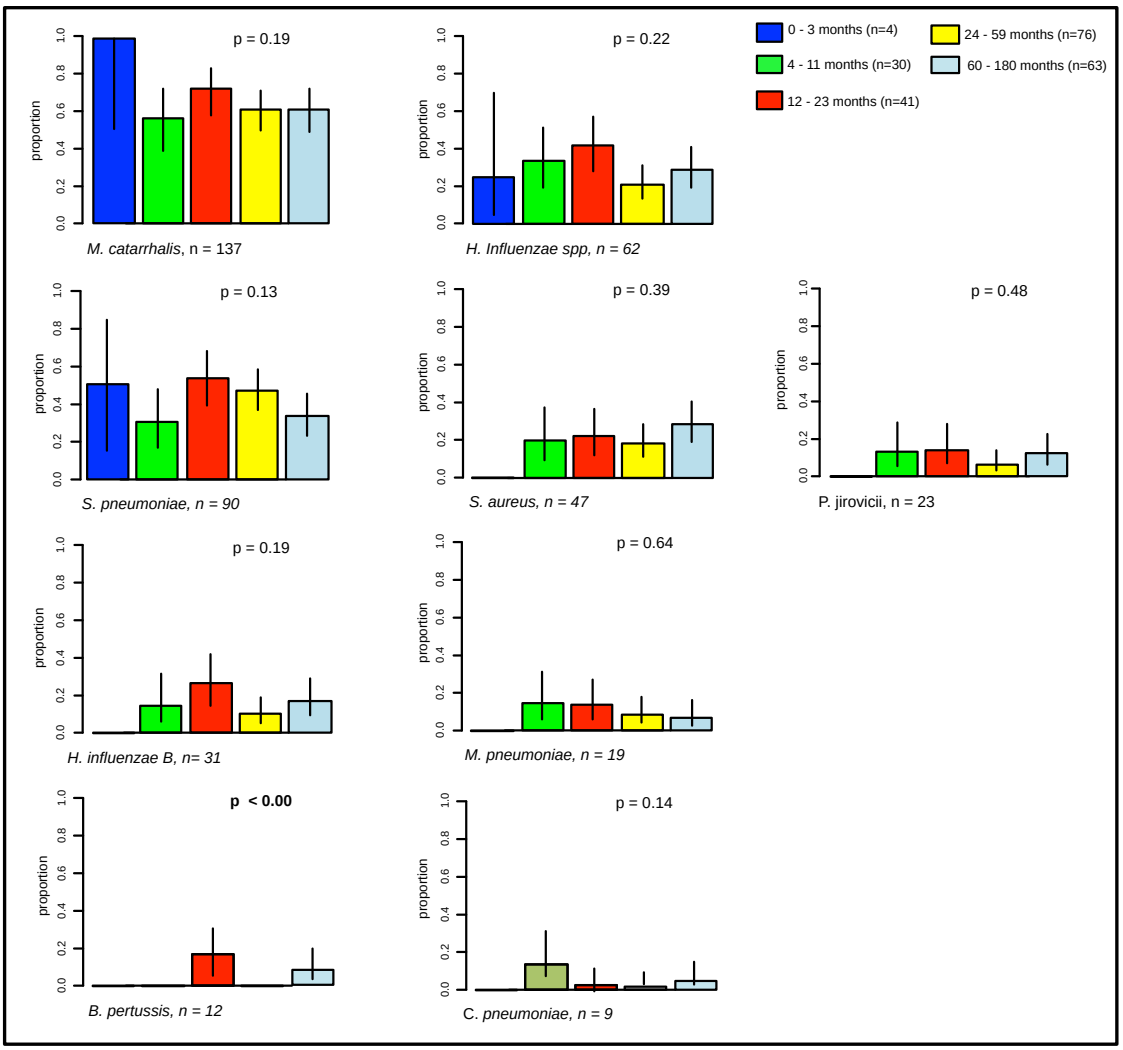

Figure 42: Age-dependent bacterial and fungal pathogen distribution among children's age groups with 95% confidence intervals. The *y*-axes represent the proportion of children in whom the bacteria/fungus was detected and the *x*-axes show age categories, as used by Tenenbaum *et al.* (19).

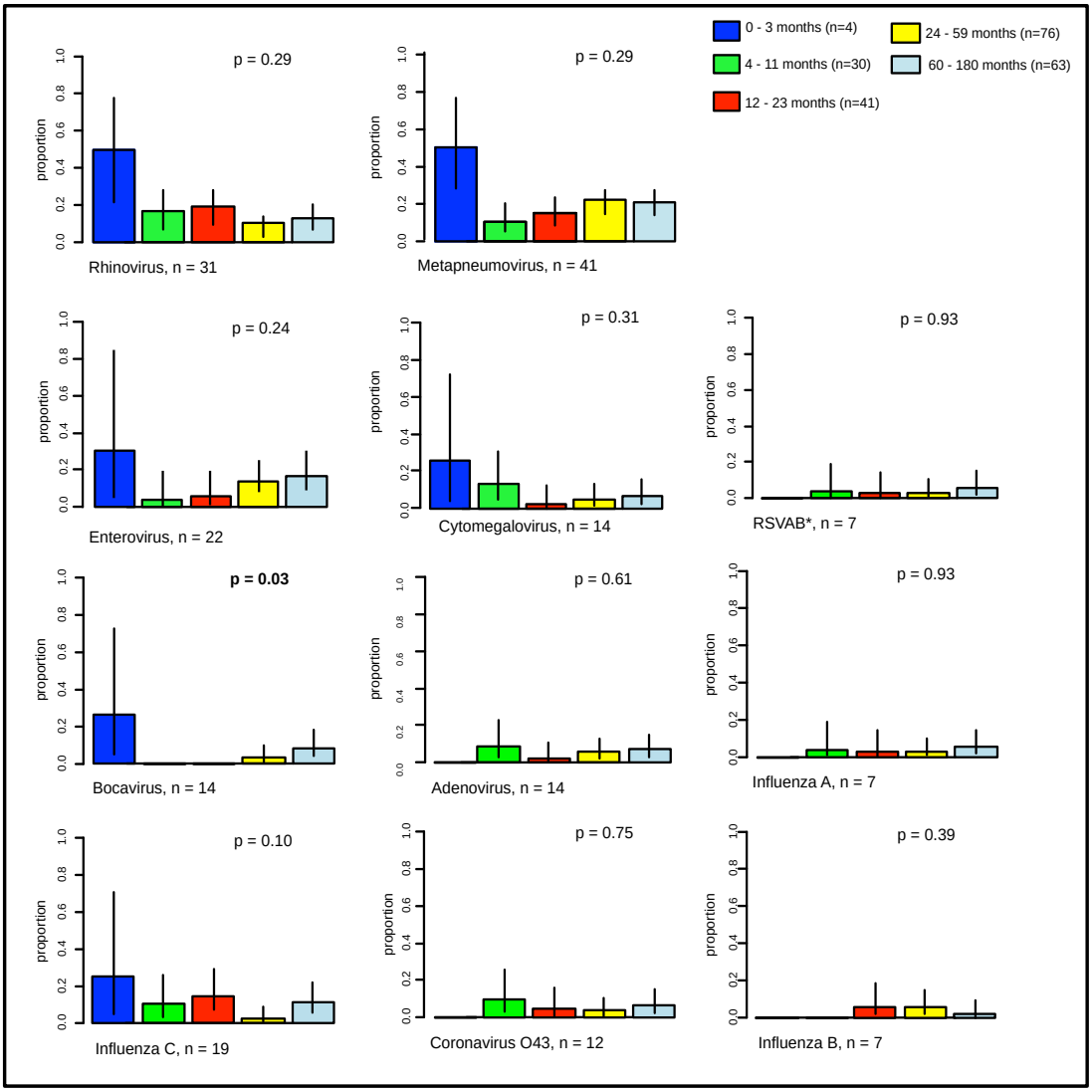

Figure 43: Age-dependent viral pathogen distribution among children's age groups with 95% confidence intervals. The *y*-axes represents the proportion of children in whom a virus was detected and the *x*-axes show age categories, as used by Tenenbaum *et al.* (19).

6.4.4 Seasonality patterns

Seasonal patterns were observed for the most prevalent viruses namely, hMPV, rhinovirus, enterovirus and influenza viruses in both 2011 and 2012. In all instances, seasonal peaks were observed over late winter and spring (figure 46). No distinct seasonal patterns for the rest of the pathogens, although numbers were small (results not shown)

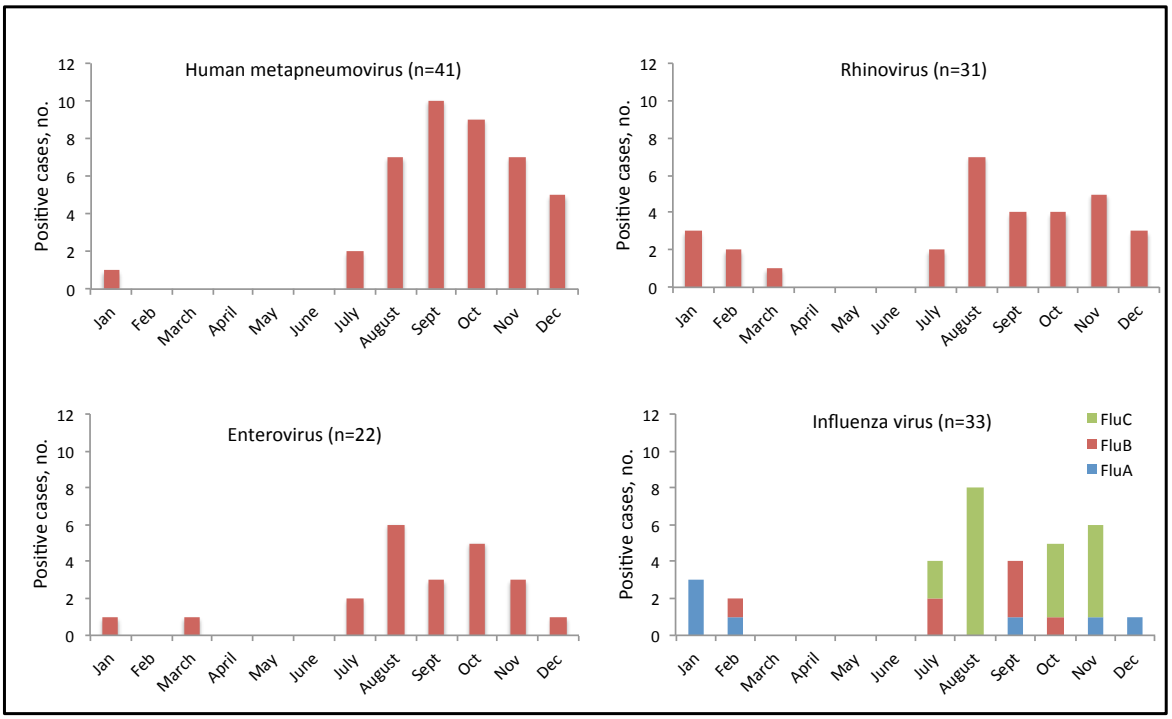

Figure 44: Seasonal distribution of the most prevalent viruses.

6.4.5 Co-infection patterns

Among the 214 nasopharyngeal samples that were collected, 97% (207/214) tested positive for at least one potential respiratory pathogen. A single bacterium was detected in 18% (39/214) of samples tested, 25% (53/214) had two bacterial species detected and 52% (111/214) had three or more bacterial species detected. A single viral pathogen was detected in 33% (71/214) of the samples tested, 22% (47/214) had two viruses detected and 5% (10/214) had three or more viruses detected (Table 24). Bacterial-viral coinfections occurred in 73% (24/33), 55% (52/95) and 56% (48/86) of the definite TB, possible TB and not TB groups respectively.

Table 23: Total number of bacterial and viral co-infections*

	Number of bacteria detected					
	0	1	2	3	>3	Number of samples
No viral infections	7	25	20	21	13	**86**
1 viral infection	2	9	18	24	18	**71**
2 viral infections	2	2	13	16	14	**47**
≥3 viral infections	0	3	2	3	2	**10**
Number of samples	**11**	**39**	**53**	**64**	**47**	**214**

**Excludes fungal infections*

A detailed overview of all possible pairs of co-occuring potential respiratory pathogens, irrespective of TB category, is shown in table 25. These co-occurences were further tested for significance (Table 6). Of note, bacterial-bacterial associations that were more common than would have been expected by chance included: *H. influenzae* type b and *M. pneumoniae* ($p = 0.04$), *M. catarrhalis* and *M. pneumoniae* ($p = 0.01$), *M. catarrhalis* and *S. pneumoniae* ($p = 0.01$), *H. influenzae spp*. and *S. pneumoniae* ($p = 0.01$), *H. influenzae* and *M. catarrhalis* ($p = 0.03$), and *B. pertussis* and *H. influenzae* type b ($p < 0.01$). Viral-viral associations co-occurring significantly more than would be expected by chance, included: bocavirus and influenza A virus ($p = 0.02$), parainfluenza 1 virus and coronavirus NL63 ($p = 0.01$) as well as hMPV and enterovirus ($p = 0.01$) (Table 6). Viral-bacterial associations co-occurring significantly more than would be expected by chance, included: hMPV and *S. pneumoniae* ($p = 0.03$), enterovirus and *H. influenzae type b* ($p = 0.03$), coronavirus O43 and *M. catarrhalis* ($p = 0.03$), influenza C virus and *M. pneumoniae* ($p = 0.02$), parainfluenza virus 3 and *M. pneumoniae* ($p = 0.01$), influenza C virus and *H. influenzae type b* ($p < 0.01$), influenza C virus and *B. pertussis* ($p < 0.01$), influenza A virus and *H. influenzae spp* ($p = 0.02$), hMPV and *H. influenzae spp* ($p = 0.02$), and cytomegalovirus and *H. influenzae spp* ($p < 0.001$). *P. jirovecii* was associated with coronavirus HKU ($p = 0.01$), *S. aureus* ($p < 0.001$) and *H. influenzae* type b ($p < 0.001$).

Table 24: Summary of all paired pathogen co-occurrence counts*

	FluA	Rh	FluB	P3	P2	P4	C29	C63	C43	HKU	RSV	CV	AV	EPV	P1	HmPV	Mp	Hbov	Sa	Cp	Sp	HiB	PCP	Mx	FluC	Hif	Bp
Influenza A (FluA)	**7**																										
Rhinovirus (Rh)	0	**31**																									
Influenza B (FluB)	0	0	**7**																								
Parainfuenza 3 (P3)	0	1	0	**2**																							
Parainfuenza 2 (P2)	0	1	0	0	**1**																						
Parainfuenza 4 (P4)	0	1	0	0	0	**2**																					
Coronavirus 229 (C229)	0	0	0	0	0	0	**1**																				
Coronavirus 63 (C63)	0	0	0	0	0	0	0	**2**																			
Coronavirus 43 (C43)	0	2	1	0	0	0	0	0	**12**																		
Coronavirus HKU	1	0	0	0	0	0	0	0	0	**2**																	
RSV A/B[a] (RSV)	0	1	0	0	0	0	0	0	0	0	**7**																
Cytomegalovirus (CV)	2	0	0	0	0	0	1	0	1	1	0	**14**															
Adenovirus (AV)	1	1	0	0	0	0	0	0	1	1	0	2	**14**														
Enterovirus (EPV)	0	6	1	0	0	1	0	0	0	0	0	1	3	**22**													
Parainfluenza 1 (P1)	0	0	0	0	0	0	0	1	1	0	0	0	0	0	**5**												
HmPV A/B[b] (HmPV)	1	6	1	1	0	0	0	0	3	0	0	4	2	9	0	**41**											
M. pneumoniae (Mp)	1	4	0	2	0	1	0	0	1	0	2	0	2	3	0	4	**19**										
Bocavirus (Hbv)	2	2	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	**8**									
S. aureus (Sa)	0	6	1	0	0	1	0	0	2	0	1	4	1	8	0	11	3	3	**47**								
C.pneumoniae (Cpn)	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	**9**							
S.pneumoniae (Spn)	5	17	3	0	0	1	0	0	7	1	1	8	5	9	1	23	7	3	22	1	**90**						
H.influenzae B (HiB)	0	4	0	0	0	1	0	0	1	0	0	2	4	7	0	5	6	0	8	1	14	**31**					
P. jirovecii (PCP)	1	4	1	0	0	0	0	1	2	2	1	2	1	2	0	4	2	1	**13**	0	9	9	**23**				
M. cartaarhalis (Mx)	5	24	6	2	0	1	0	2	11	1	5	10	6	18	2	31	17	2	28	4	75	19	14	**137**			
Influenza C(FluC)	0	2	0	0	0	0	0	0	2	0	2	0	1	2	2	2	5	1	0	2	3	8	2	13	**19**		
H. influenzae spp (Hif)	5	9	3	1	0	0	0	0	4	1	1	9	5	4	1	18	5	2	8	3	35	7	6	46	0	**62**	
B. pertussis (Bp)	0	2	1	0	0	0	0	0	1	0	0	0	0	3	0	1	2	0	0	1	3	7	0	10	7	4	**12**

**The pathogen abbreviation names in the top row are derived from the first column. The diagonal line represents the total count of each pathogen detected.* [a]*RSV A/B = Respiratory Syncytial virus A and B,* [b]HmPV A/B = Human Metapneumovirus A and B

Table 25: Summary of achieved significance level for all paired pathogen co-occurrences*

	FluA	Rh	FluB	P3	P2	P4	C29	C63	C43	HKU	RSV	CV	AV	EPV	P1	HmPV	Mp	Hbov	Sa	Cp	Sp	HiB	PCP	Mx	FluC	Hif
Rhinovirus (Rh)	1																									
Influenza B (FluB)	1	1																								
Parainfuenza 3 (P3)	1	0.27	1																							
Parainfuenza 2 (P2)	1	0.14	1	1																						
Parainfuenza 4 (P4)	1	0.27	1	1	1																					
Coronavirus 229 (C229)	1	1	1	1	1	1																				
Coronavirus 63 (C63)	1	1	1	1	1	1	1																			
Coronavirus 43 (C43)	1	0.56	0.33	1	1	1	1	1																		
Coronavirus HKU	0.07	1	1	1	1	1	1	1	1																	
RSV A/B[a] (RSV)	1	0.66	1	1	1	1	1	1	1	1																
Cytomegalovirus (CV)	0.07	1	1	1	1	1	0.07	1	0.57	0.12	1															
Adenovirus (AV)	0.38	0.90	1	1	1	1	1	1	0.56	0.13	1	0.23														
Enterovirus (EPV)	1	0.08	0.54	1	1	0.20	1	1	1	1	1	0.80	0.16													
Parainfluenza 1 (P1)	1	1	1	1	1	1	1	**0.04**	0.25	1	1	1	1	1												
HmPV A/B[b] (HmPV)	0.78	0.57	0.78	0.3	1	1	1	1	0.41	1	1	0.27	0.79	**0.01**	1											
M. pneumoniae (Mp)	0.48	0.29	1	**0.01**	1	0.17	1	1	0.68	1	0.12	1	0.36	0.31	1	0.51										
Bocavirus (Hbv)	**0.02**	0.33	1	1	1	1	1	1	1	0.08	0.24	1	0.42	0.58	1	0.82	1									
S. aureus (Sa)	1	0.72	0.83	1	1	0.39	1	1	0.79	1	0.82	0.37	0.97	0.08	1	0.26	0.83	0.25								
C.pneumoniae (Cp)	1	0.76	1	1	1	1	1	1	1	1	0.26	0.47	0.46	0.63	0.19	1	0.58	0.29	0.89							
S.pneumoniae (Sp)	0.11	0.09	0.62	1	1	0.67	1	1	0.19	0.67	0.98	0.18	0.77	0.63	0.94	**0.03**	0.76	0.73	0.28	0.99						
H.influenzae B (HiB)	1	0.69	1	1	1	0.27	1	1	0.85	1	1	0.63	0.12	**0.02**	1	0.75	**0.04**	1	0.36	0.76	0.43					
P. jirovecii	0.55	0.44	0.55	1	1	1	1	0.21	0.38	**0.01**	0.55	0.46	0.81	0.72	1	0.68	0.63	0.60	**0.00**	1	0.70	**0.00**				
M. cartaarhalis (Mx)	0.51	0.07	0.21	0.42	1	0.88	1	0.41	**0.03**	0.87	0.51	0.38	0.97	**0.05**	0.95	0.06	**0.01**	1	0.81	0.94	**0.00**	0.71	0.72			
Influenza C(FluC)	1	0.79	1	1	1	1	1	1	0.30	1	0.12	1	0.75	0.61	0.06	0.91	**0.02**	0.53	1	0.19	1	**0.00**	0.63	0.45		
H. influenzae spp (Hif)	**0.02**	0.57	0.32	0.49	1	1	1	1	0.47	0.50	0.91	**0.00**	0.39	0.92	0.82	**0.02**	0.69	0.73	0.99	0.52	**0.01**	0.86	0.71	**0.03**	1	
B. pertussis (Bd)	1	0.54	0.33	1	1	1	1	1	0.51	1	1	1	1	0.12	1	0.93	0.29	1	1	0.41	0.94	**0.00**	1	0.13	**0.00**	0.48

**The pathogen abbreviation names in the top row are derived from the first column. The diagonal line represents comparison to self which empirically has an achieved level of significance (ASL) of p=1, the table reports ASL for all concurrent pathogen pairs irrespective of the TB category. Significant concurrent relationships are highlighted in bold red. [a]RSV A/B = Respiratory Syncytial virus A and B,* [b]HmPV A/B = Human Metapneumovirus A and B

6.4.6 Concurrent respiratory pathogens by TB category

No clear differences in the patterns of respiratory pathogens detected in this cohort by TB status were found, either by direct comparison (Table 23) or using LDA, where there was no clear separation of groups, figure 47. However, children with definite TB tended to have *C. pneumoniae*, coronavirus 043, *S. pneumoniae,* rhinovirus, parainfluenza virus 1, adenovirus or *M. catarrhalis,* figure 47. The non-TB group had a variety of respiratory pathogens including cytomegalovirus, influenza B, *S. aureus,* parainfluenza virus 2, *H. influenzae spp.,* RSV A/B and *P. jirovecii*, figure 47. The possible TB group was largely comprised of a mixture of respiratory pathogens from either definite or non-TB group; therefore, the three groups could not be optimally separated.

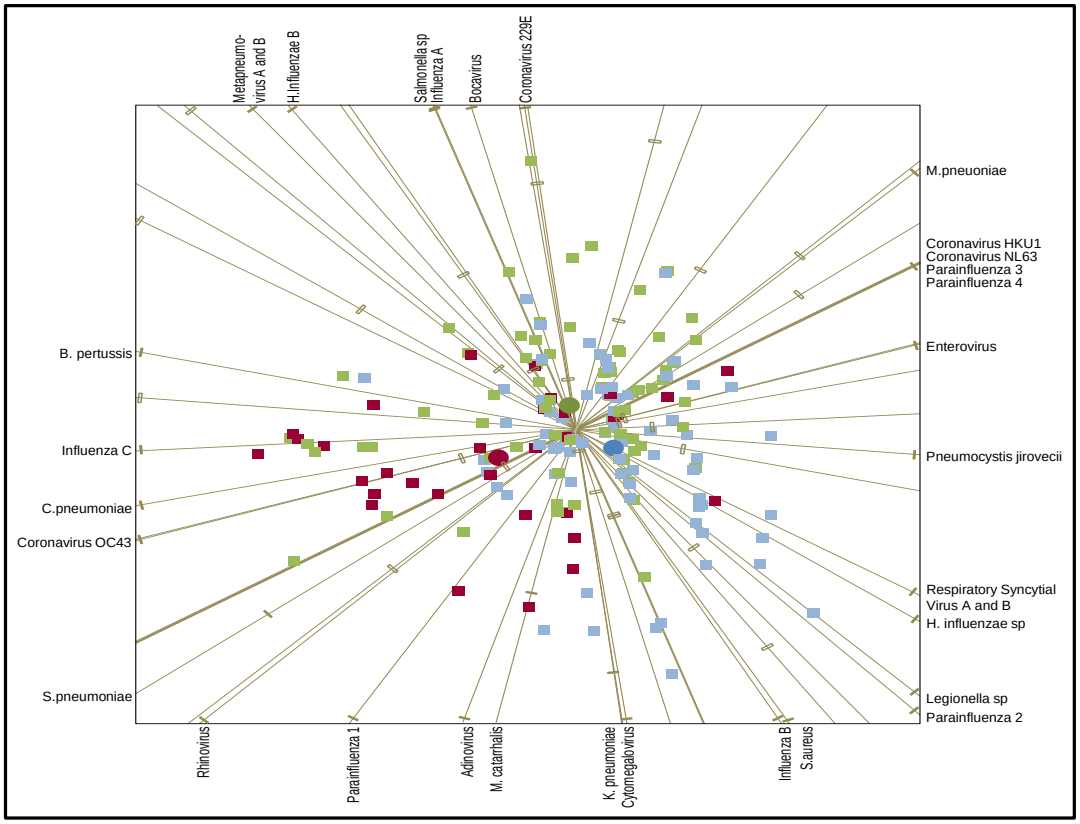

Figure 45: Canonical Variate Analysis (CVA) biplot for graphical visualisation of Linear Discriminant Analysis (LDA). This demonstrates the presence or absence of respiratory

pathogens in the definite TB (red squares), not TB (blue squares) and possible TB (green squares) groups.

In further analyses, the possible TB cases were excluded and only definite TB and not TB groups were considered. By doing this, the CVA biplot was reduced to one dimension as seen by the projection of all pathogens onto a line (figure 48). There was a large amount of overlap between the two groups.

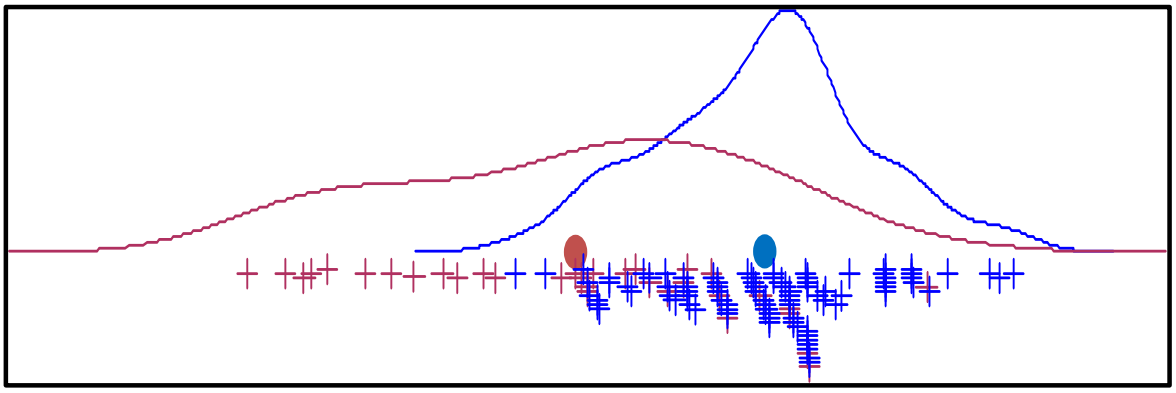

Figure 46: Canonical Variate Analysis (CVA) biplot depicting the spread of respiratory pathogens in the definite TB (red line) and not TB (blue line) groups only. Observations under each group are denoted by "+" signs and the median of each group by the red and blue ovals.

Given the variance detected from the one dimensional LDA, we further explored the use of a QDA approach (which accounts for different variances between the definite TB and non-TB groups) to determine which set of pathogens optimally discriminates between the TB groups, figure 49 below. HMPV, *C. pneumoniae*, coronavirus 043, influenza C virus, rhinovirus and cytomegalovirus grouped more towards the definite TB. In contrast, *M. pneumoniae*, *H. influenzae spp.*, *P. jirovecii*, enterovirus, influenza B virus and RSV A/B grouped toward the non-TB category. *S. pneumoniae* did not cluster to any particular TB category.

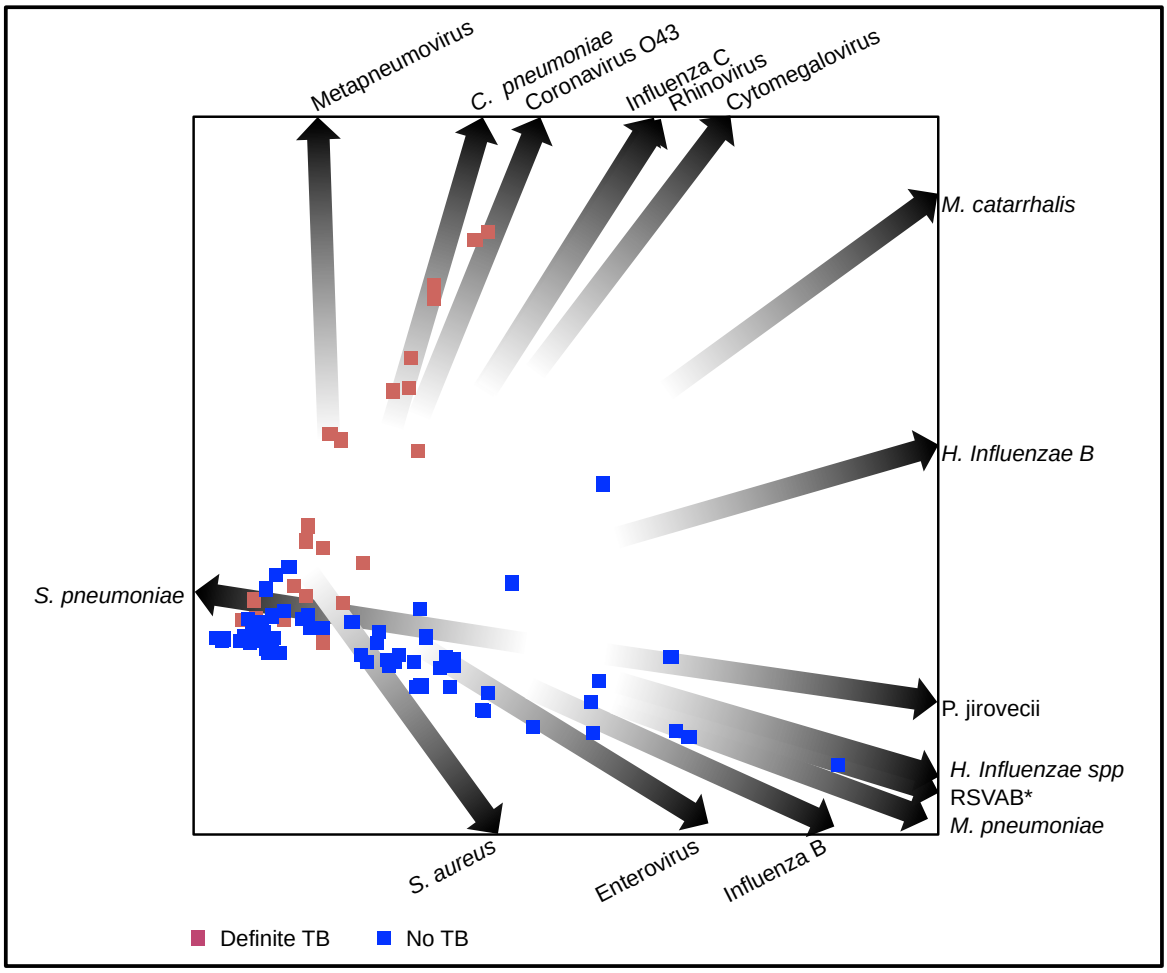

Figure 47: Quadratic Discriminant Analysis (QDA) biplot for graphical visualisation of pathogen classification. Clustering of respiratory pathogens in the definite TB (red squares) and non TB (blue squares) groups is shown. Since the pathogens were recorded on a presence/absence scale rather than quantitative results, the biplot axes are replaced with arrows indicating the direction of presence of the pathogens. *RSVAB = Respiratory Syncytial virus A and B.

6.5 DISCUSSION

This study describes for the first time, the diversity of respiratory pathogens, other than MTB, detected in children presenting to hospital with suspected TB in a high TB and HIV endemic setting. Of the children admitted to hospital with suspected pulmonary TB, 16% (34/214) had microbiologically culture confirmed TB, while 40% (86/214) had a negative MTB culture and were therefore classified as having possible TB based on

clinical and radiological findings. The observed culture-confirmed rate of TB is consistent with that of McNally *et al*. (5) but was high when compared to the reported 5 - 8% from studies in similarly TB endemic settings amongst children presenting with acute respiratory infections (8,48–51). These differences may be attributed to differences in the study populations, enrollment criteria and diagnostic tests utilised across the studies. The non-specific clinical presentation of PTB in infants and young children and the potential contribution of co-infections (bacterial, viral and fungal) highlights a need for improved diagnosis of PTB and identification of respiratory tract pathogens other than MTB in children with suspected PTB.

There was no clear evidence of age-dependent infection patterns in this cohort except for bocavirus and *B. pertussis* which were more commonly detected in children younger than three months ($p = 0.03$) and those between 12 months and 23 months old respectively ($p < 0.01$). This is in contrast with a study by Fry *et al*. who found that Thai children below the age of one year were less likely to be infected with bocavirus compared with children between one and four years of age (13% vs. 70%) (52). However, the authors also found a low prevalence of bocavirus infections 4.5% (53/1168), which was similar to the prevalence found in our study. Although a number of other viruses were mostly detected in children younger than three months, the small sample size limits our ability to draw conclusions from these results. According to a population-based study in Kilifi, Kenya, RSV was more prevalent among infants, compared to children 12 - 59 months old (33% [45/137] vs 17% [24/117], p<0.01) (53). We did not observe any differences in the prevalence of RSV infection across any age groups, however numbers were very small.

Overall, at least one potential respiratory pathogen was detected in 97% of the respiratory specimens collected from the children in our cohort. Similar detection rates

have been shown elsewhere amongst children hospitalised with lower respiratory infections (LRTI) (34,54). NP colonisation/infection by a single bacterial species was fairly uncommon compared to colonisation by more than three bacterial species (18% vs 52%, $p < 0.01$). The most common colonising/infecting bacteria detected in this study include *M. catarrhalis*, *S. pneumoniae*, *H. influenzae*, and *S. aureus* (64%, 42%, 29% and 22%, respectively). These findings are in line with the nasopharyngeal colonisation rates described elsewhere in children with lower respiratory tract infections (17,19,20,55).

The prevalence of viral infection in this cohort is lower than that previously reported in children with acute lower respiratory infections (ALRI) (6,7,9,10,13,56). This may be explained by the enrolment criteria that encompassed children with symptoms suggestive of PTB and with a long duration of cough, whereas most of the other studies focused on children with acute symptoms, which are more likely to be of viral origin. The most commonly detected viruses were hMPV (19%), rhinovirus (15%), influenza virus (9%), adenovirus (7%), cytomegalovirus (7%) and coronavirus O43 (6%). Similar viral detection rates have been reported elsewhere amongst children hospitalised for either community acquired pneumonia (CAP) or ALRI (6,7,15,16,18,57). The reported range of detection rates for each of the above viruses were: hMPV (1 - 13%), rhinovirus (20 - 45%), influenza virus (4 - 14%), adenovirus (6 - 13%) and coronavirus O43 (1 - 7%) (6,7,16,18,34).

Care needs to be taken in the interpretation of molecular detection of rhinovirus, cytomegalovirus, coronavirus, enterovirus and bocavirus in respiratory specimens, since these viruses may be detected in children without lower respiratory tract infections (34,57–61). However, detection of RSV, hMPV, adenovirus, influenza virus

and some coronavirus subtypes is typically associated with pneumonia than is detection of other respiratory viruses (52,58,62–65).

Our understanding of microbial interactions is, in part, constrained by the tools that allow us to adequately visualise such interactions without reducing vital data dimensions. Using a novel methodology, QDA, we have shown that the only bacterial species that clustered more towards children with microbiologically confirmed TB was *C. pneumoniae*. The significance of detecting other potential bacterial pathogens in children presenting with prolonged symptoms suggestive of TB is unclear. One possibility is that the relative immune suppression associated with TB may render the host susceptible to a bacterial infection, or alternatively, that reactivation of latent TB may occur following intracellular bacterial infection (66-69).

Viruses that clustered with the TB group include hMPV, coronavirus O43, influenza C virus, rhinovirus and cytomegalovirus. MTB has a slow replication rate. Therefore, acute viral infections in patients with undiagnosed TB disease may precipitate an acute exacerbation of symptoms severe enough to require hospital admission and expeditious testing. Anecdotal reports from influenza pandemics have reported high mortality rates in individuals with an underlying TB disease (70-73). A recent South African study has shown an increased risk of death in individuals with TB-influenza A virus co-infection (adjusted relative risk ratio [aRRR] 6.1, 95% confidence interval [CI] 1.6 – 23.4) compared to TB infection alone (74). In contrast, de Paus *et al.* did not find a correlation between the seroprevalence of influenza antibodies and the development of clinically active TB in an Indonesian cohort. They did however show an association between elevated antibody titers against influenza A and the clinical stage of TB lung disease suggesting recent re-infection with influenza precedes clinical manifestation of TB (75).

Although rhinoviruses, which also clustered in the TB group, are most commonly associated with mild upper respiratory tract infections (59,60,76), infections of lower airways do occur (58,77,78). These include bronchiolitis, pneumonia, chronic bronchitis and the development of asthma (57,58,77,78). Some rhinovirus genotypes, specifically subtype C, correlate with ALRI disease severity (14,58,59,61,79). A more comprehensive study of rhinovirus genotypes in this cohort would be of interest (14,58,59,61,79).

On the other hand, respiratory pathogens that clustered with the non-TB group included *P. jirovecii*, *H. influenzae spp.,* RSV, *M. pneumoniae*, influenza B virus and enteroviruses. While disease causation cannot be inferred by the mere presence of bacteria in the nasopharynx, the presence of either *M. pneumoniae*, *C. pneumoniae* and *B. pertussis* has been shown to be associated with ALRI (56,80–82). The incidence of *M. pneumoniae* and *C. pneumoniae* was generally low (10% and 4% respectively) and appeared to be more common in children younger than two years of age. Similar detection rates have been reported in children hospitalised with ALRI (9,83,84). Bezerra *et al*. (9) examined the aetiological causes of ALRI of varying clinical severity among 407 hospitalised children in Brazil. They showed that *M. pneumoniae* was more likely to be detected in children hospitalised with pneumonia than ambulatory bronchiolitis cases (13.8% vs. 7.1% respectively, $p = 0.04$). In contrast, Tsolia *et al*. (7) and Principi *et al*. (85) have reported higher rates of *M. pneumoniae* infection in children hospitalised with ALRI (36% and 38%, respectively). The relatively low prevalence of *M. pneumoniae* in our study may be related to the enrollment criteria, and relatively long duration of symptoms amongst children enrolled.

The prevalence of *B. pertussis* in the present study was 6%, and was highest in children between 12 and 23 months. Similar infection frequencies of 1 – 9% have been observed

in other settings in children with ALRI, 10 – 20 years post transition from whole-cell vaccines (WCV) to acellular vaccines (ACV) (86–89). Persistence and resurgence of pertussis has been attributed to waning immunity and/or pathogen adaptation in response to vaccination (Reviewed by (90)).

RSV infections were less frequently encountered in our cohort in contrast to other reported prevalences (3% vs 10 - 40%) amongst children with ALRI (9,10,13,16,18,91–93). The lower detection rates of RSV in our cohort may be attributed to enrollment criteria and long symptom duration. The recent finding of novel respiratory syncytial virus A genotype ON1 (originally characterised in Ontario, Canada) in a Cape Town paediatric hospital further highlights the need for continued surveillance of RSV infections (94). Our study did not attempt to determine RSV genotype.

Some pathogens were not associated with any of the two TB categories. These included *M. catarrhalis*, *H. influenzae* type b and *S. pneumoniae*. A randomised controlled trial that assessed the effect of a 9-valent pneumococcal conjugate vaccine (PCV9) in South African children showed decreased rates of culture-confirmed and clinically diagnosed TB in PCV9 recipients compared with placebo recipients (relative risk reduction 43% [95% CI, 11.5% - 63.8%], $p = 0.0114$) (8). This suggests an intriguing possibility that *S. pneumoniae* coinfection might increase children's vulnerability to develop active TB. Although common in our cohort, *S. pneumoniae* was only detected in 20% (18/90) of the children with TB and did not cluster together with pathogens in the TB or non-TB groups.

We recognise that from a diagnostic perspective, the most appropriate specimen to use would be a representative sample from the disease affected tissue, in this case lung aspirates or good quality sputum. Induced sputum was not available for analysis as it was used entirely for TB investigations. The present study did not study respiratory

pathogens in an appropriately matched control group. Therefore, it was not possible to draw inferences whether the detected pathogens contributed to the respiratory illness in these children with the exception of specific pathogens such as *B. pertussis* (89,90,95), *M. pneumoniae*, *C. pneumoniae* (56,80–83), hMPV, influenza A/B viruses and RSV (76,96). These pathogens have been strongly associated with disease in other studies. Studies that clearly differentiate between infection and colonisation are essential for advancing therapeutic strategies and prioritising diagnostic efforts.

In conclusion, this study described the detection of multiple respiratory pathogens in the nasopharynx of children hospitalised with suspected TB. Whilst there was no clear separation between the pathogens present in the airways of children with and without PTB, *C. pneumoniae*, hMPV, coronavirus O43, influenza C virus, rhinovirus and cytomegalovirus were associated with PTB. On the other hand, *P. jirovecii*, *H. influenzae spp*, RSV, *M. pneumoniae*, influenza B virus and enteroviruses were more common amongst children without TB. This pilot work may signal broader differences in the microbial ecology of the upper respiratory tract of these children, which warrants further study.

6.6 REFERENCES

1. Liu L, Johnson HL, Cousens S, Perin J, Scott S, et al. (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 379: 2151–2161.
2. Lin PL, Flynn JL (2010) Understanding latent tuberculosis: a moving target. J Immunol 185: 15–22.
3. Oliwa JN, Karumbi JM, Marais BJ, Madhi SA, Graham SM (2015) Tuberculosis as a cause or comorbidity of childhood pneumonia in tuberculosis-endemic areas: a systematic review. Lancet Respir Med 3: 235–243.
4. Connell TG, Zar HJ, Nicol MP (2011) Advances in the diagnosis of pulmonary tuberculosis in HIV-infected and HIV-uninfected children. J Infect Dis 204 Suppl : S1151–S1158.
5. McNally LM, Jeena PM, Gajee K, Thula SA, Sturm AW, et al. (2007) Effect of age, polymicrobial disease, and maternal HIV status on treatment response and cause of severe pneumonia in South African children: a prospective descriptive study. Lancet 369: 1440–1451.
6. Cevey-Macherel M, Galetto-Lacour A, Gervaix A, Siegrist C-A, Bille J, et al. (2009) Etiology of community-acquired pneumonia in hospitalized children based on WHO clinical guidelines. Eur J Pediatr 168: 1429–1436.
7. Tsolia MN, Psarras S, Bossios A, Audi H, Paldanius M, et al. (2004) Etiology of community-acquired pneumonia in hospitalized school-age children: evidence for high prevalence of viral infections. Clin Infect Dis 39: 681–686.
8. Moore DP, Klugman KP, Madhi SA (2010) Role of *Streptococcus pneumoniae* in hospitalization for acute community-acquired pneumonia associated with culture-confirmed Mycobacterium tuberculosis in children: a pneumococcal conjugate vaccine probe study. Pediatr Infect Dis J 29: 1099–04.
9. Bezerra PGM, Britto MCA, Correia JB, Duarte M do CMB, Fonceca AM, et al. (2011) Viral and atypical bacterial detection in acute respiratory infection in children under five years. PLoS One 6: e18928.
10. Nicol MP, Workman L, Isaacs W, Munro J, Black F, et al. (2011) Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study. Lancet Infect Dis 11: 819–824
11. Bosis S, Esposito S, Osterhaus ADME, Tremolati E, Begliatti E, et al. (2008) Association between high nasopharyngeal viral load and disease severity in children with human metapneumovirus infection. J Clin Virol 42: 286–290.
12. Brieu N, Guyon G, Rodière M, Segondy M, Foulongne V (2008) Human bocavirus infection in children with respiratory tract disease. Pediatr Infect Dis J 27: 969–973.
13. Cilla G, Oñate E, Perez-Yarza EG, Montes M, Vicente D, et al. (2008) Viruses in community-acquired pneumonia in children aged less than 3 years old: High rate

of viral coinfection. J Med Virol 80: 1843–1849.

14. Ferguson PE, Gilroy NM, Faux CE, Mackay IM, Sloots TP, et al. (2013) Human rhinovirus C in adult haematopoietic stem cell transplant recipients with respiratory illness. J Clin Virol 56: 255–259.

15. Honkinen M, Lahti E, Österback R, Ruuskanen O, Waris M (2012) Viruses and bacteria in sputum samples of children with community-acquired pneumonia. Clin Microbiol Infect 18: 300–307.

16. Hoffmann J, Rabezanahary H, Randriamarotia M, Ratsimbasoa A, Najjar J, et al. (2012) Viral and atypical bacterial etiology of acute respiratory infections in children under 5 years old living in a rural tropical area of Madagascar. PLoS One 7: e43666.

17. Pettigrew MM, Gent JF, Revai K, Patel JA, Chonmaitree T (2008) Microbial interactions during upper respiratory tract infections. Emerg Infect Dis 14: 1584–1591.

18. Pretorius MA, Madhi SA, Cohen C, Naidoo D, Groome M, et al. (2012) Respiratory viral coinfections identified by a 10-plex real-time reverse-transcription polymerase chain reaction assay in patients hospitalized with severe acute respiratory illness--South Africa, 2009-2010. J Infect Dis 206 Suppl : S159–S165.

19. Tenenbaum T, Franz A, Neuhausen N, Willems R, Brade J, et al. (2012) Clinical characteristics of children with lower respiratory tract infections are dependent on the carriage of specific pathogens in the nasopharynx. Eur J Clin Microbiol Infect Dis 31: 3173–3182.

20. van den Bergh MR, Biesbroek G, Rossen JWA, de Steenhuijsen Piters WAA, Bosch AATM, et al. (2012) Associations between pathogens in the upper respiratory tract of young children: interplay between viruses and bacteria. PLoS One 7: e47711.

21. Arens MQ, Buller RS, Rankin A, Mason S, Whetsell A, et al. (2010) Comparison of the Eragen Multi-Code Respiratory Virus Panel with conventional viral testing and real-time multiplex PCR assays for detection of respiratory viruses. J Clin Microbiol 48: 2387–2395.

22. Bibby DF, McElarney I, Breuer J, Clark DA (2011) Comparative evaluation of the Seegene Seeplex RV15 and real-time PCR for respiratory virus detection. J Med Virol 83: 1469–1475.

23. Bierbaum S, Forster J, Berner R, Rücker G, Rohde G, et al. (2014) Detection of respiratory viruses using a multiplex real-time PCR assay in Germany, 2009/10. Arch Virol 159: 669–676.

24. Brittain-Long R, Nord S, Olofsson S, Westin J, Anderson L-M, et al. (2008) Multiplex real-time PCR for detection of respiratory tract infections. J Clin Virol 41: 53–56.

25. Brittain-Long R, Westin J, Olofsson S, Lindh M, Andersson L-M (2010) Prospective evaluation of a novel multiplex real-time PCR assay for detection of fifteen respiratory pathogens-duration of symptoms significantly affects detection rate. J Clin Virol 47: 263–267.

26. Gadsby NJ, Hardie A, Claas ECJ, Templeton KE (2010) Comparison of the Luminex Respiratory Virus Panel fast assay with in-house real-time PCR for respiratory viral infection diagnosis. J Clin Microbiol 48: 2213–2216.

27. Gibson UE, Heid CA, Williams PM (1996) A novel method for real time quantitative RT-PCR. Genome Res 6: 995–1001.

28. Rand KH, Rampersaud H, Houck HJ (2011) Comparison of Two Multiplex Methods for Detection of Respiratory Viruses: FilmArray RP and xTAG RVP. J Clin Microbiol 49: 2449–2453.

29. Sakthivel SK, Whitaker B, Lu X, Oliveira DBL, Stockman LJ, et al. (2012) Comparison of fast-track diagnostics respiratory pathogens multiplex real-time RT-PCR assay with in-house singleplex assays for comprehensive detection of human respiratory viruses. J Virol Methods 185: 259–266.

30. Caliendo AM (2011) Multiplex PCR and emerging technologies for the detection of respiratory pathogens. Clin Infect Dis 52 Suppl 4: S326–S330.

31. Mahony J, Chong S, Merante F, Yaghoubian S, Sinha T, et al. (2007) Development of a respiratory virus panel test for detection of twenty human respiratory viruses by use of multiplex PCR and a fluid microbead-based assay. J Clin Microbiol 45: 2965–2970.

32. Lamson D, Renwick N, Kapoor V, Liu Z, Palacios G, et al. (2006) MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004-2005. J Infect Dis 194: 1398–1402.

33. Gunson RN, Bennett S, Maclean A, Carman WF (2008) Using multiplex real time PCR in order to streamline a routine diagnostic service. J Clin Virol 43: 372–375.

34. Mermond S, Zurawski V, D'Ortenzio E, Driscoll AJ, DeLuca AN, et al. (2012) Lower respiratory infections among hospitalized children in New Caledonia: a pilot study for the Pneumonia Etiology Research for Child Health project. Clin Infect Dis 54 Suppl 2: S180–S189.

35. Anderson TP, Werno AM, Barratt K, Mahagamasekera P, Murdoch DR, et al. (2013) Comparison of four multiplex PCR assays for the detection of viral pathogens in respiratory specimens. J Virol Methods 191: 118–121.

36. Lockman S, Hone N, Kenyon TA, Mwasekaga M, Villauthapillai M, et al. (2003) Etiology of pulmonary infections in predominantly HIV-infected adults with suspected tuberculosis, Botswana. Int J Tuberc Lung Dis 7: 714–723.

37. Nicol MP, Workman L, Isaacs W, Munro J, Black F, et al. (2011) Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study. Lancet Infect Dis 11: 819–824.

38. WHO (2007) WHO Case Definitions of HIV for Surveillance and Revised Clinical Staging and Immunological Classification of HIV-Related Disease in Adults and Children.

39. Nicol MP, Allen V, Workman L, Isaacs W, Munro J, et al. (2014) Urine lipoarabinomannan testing for diagnosis of pulmonary tuberculosis in children: a

prospective study. Lancet Glob Heal 2: e278–e284.

40. National Department of Health South Africa. (2014) National Tuberculosis Management Guidelines 2014.

41. Dube FS, Kaba M, Whittaker E, Zar HJ, Nicol MP (2013) Detection of *Streptococcus pneumoniae* from Different Types of Nasopharyngeal Swabs in Children. PLoS One 8: e68097.

42. Zar HJ, Tannenbaum E, Apolles P, Roux P, Hanslo D, et al. (2000) Sputum induction for the diagnosis of pulmonary tuberculosis in infants and young children in an urban setting in South Africa. Arch Dis Child 82: 305–308.

43. Claassen S, du Toit E, Kaba M, Moodley C, Zar HJ, et al. (2013) A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. J Microbiol Methods 94: 103–110.

44. R Core Team (2014) R: A language and environment for statistical computing. Available: http://www.r-project.org.

45. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B 57: 289–300.

46. Efron B, Tibshirani R (1993) An Introduction to the Bootstrap. New York: Chapman and Hall.

47. Gardner-Lubbe S, Dube FS (2015) Visualisation of quadratic discriminant analysis and its application in exploration of microbial interactions. BioData Min 8: 8.

48. Nantongo JM, Wobudeya E, Mupere E, Joloba M, Ssengooba W, et al. (2013) High incidence of pulmonary tuberculosis in children admitted with severe pneumonia in Uganda. BMC Pediatr 13: 16.

49. Chisti MJ, Graham SM, Duke T, Ahmed T, Ashraf H, et al. (2014) A prospective study of the prevalence of tuberculosis and bacteraemia in Bangladeshi children with severe malnutrition and pneumonia including an evaluation of Xpert MTB/RIF assay. PLoS One 9: e93776.

50. Madhi SA, Petersen K, Madhi A, Khoosal M, Klugman KP (2000) Increased disease burden and antibiotic resistance of bacteria causing severe community-acquired lower respiratory tract infections in human immunodeficiency virus type 1-infected children. Clin Infect Dis 31: 170–176.

51. Zar HJ, Hanslo D, Tannenbaum E, Klein M, Argent A, et al. (2001) Aetiology and outcome of pneumonia in human immunodeficiency virus-infected children hospitalized in South Africa. Acta Paediatr 90: 119–125.

52. Fry AM, Lu X, Chittaganpitch M, Peret T, Fischer J, et al. (2007) Human bocavirus: a novel parvovirus epidemiologically associated with pneumonia requiring hospitalization in Thailand. J Infect Dis 195: 1038–1045.

53. Feikin DR, Njenga MK, Bigogo G, Aura B, Aol G, et al. (2013) Viral and bacterial causes of severe acute respiratory illness among children aged less than 5 years in a high malaria prevalence area of western Kenya, 2007-2010. Pediatr Infect Dis J 32: e14–e19.

54. Feikin DR, Kagucia EW, Loo JD, Link-Gelles R, Puhan MA, et al. (2013) Serotype-

specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: a pooled analysis of multiple surveillance sites. PLoS Med 10: e1001517.

55. Wolter N, Tempia S, Cohen C, Madhi SA, Venter M, et al. (2014) High nasopharyngeal pneumococcal density, increased by viral coinfection, is associated with invasive pneumococcal pneumonia. J Infect Dis 210: 1649–1657.

56. Juvén T, Mertsola J, Waris M, Leinonen M, Meurman O, et al. (2000) Etiology of community-acquired pneumonia in 254 hospitalized children. Pediatr Infect Dis J 19: 293–298.

57. Renwick N, Schweiger B, Kapoor V, Liu Z, Villari J, et al. (2007) A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. J Infect Dis 196: 1754–1760.

58. Papadopoulos NG, Bates PJ, Bardin PG, Papi A, Leir SH, et al. (2000) Rhinoviruses infect the lower airways. J Infect Dis 181: 1875–1884.

59. Mackay IM (2008) Human rhinoviruses: the cold wars resume. J Clin Virol 42: 297–320.

60. Jartti T, Jartti L, Peltola V, Waris M, Ruuskanen O (2008) Identification of respiratory viruses in asymptomatic subjects: asymptomatic respiratory viral infections. Pediatr Infect Dis J 27: 1103–1107.

61. Broberg E, Niemelä J, Lahti E, Hyypiä T, Ruuskanen O, et al. (2011) Human rhinovirus C--associated severe pneumonia in a neonate. J Clin Virol 51: 79–82.

62. Berkley JA, Munywoki P, Ngama M, Kazungu S, Abwao J, et al. (2010) Viral etiology of severe pneumonia among Kenyan infants and children. JAMA 303: 2051–2057.

63. Hammitt LL, Kazungu S, Morpeth SC, Gibson DG, Mvera B, et al. (2012) A preliminary study of pneumonia etiology among hospitalized children in Kenya. Clin Infect Dis 54 Suppl 2: S190–S199.

64. Mathisen M, Strand TA, Valentiner-Branth P, Chandyo RK, Basnet S, et al. (2010) Respiratory viruses in nepalese children with and without pneumonia: a case-control study. Pediatr Infect Dis J 29: 731–735.

65. Nokes DJ, Okiro EA, Ngama M, Ochola R, White LJ, et al. (2008) Respiratory syncytial virus infection and disease in infants and young children observed from birth in Kilifi District, Kenya. Clin Infect Dis 46: 50–57.

66. Dudani R, Murali-Krishna K, Krishnan L, Sad S (2008) IFN-gamma induces the erosion of preexisting CD8 T cell memory during infection with a heterologous intracellular bacterium. J Immunol 181: 1700–1709.

67. Barkman C, Martner A, Hessle C, Wold AE Soluble bacterial constituents down-regulate secretion of IL-12 in response to intact Gram-positive bacteria. Microbes Infect 10: 1484–1493.

68. Jo E-K (2008) Mycobacterial interaction with innate receptors: TLRs, C-type lectins, and NLRs. Curr Opin Infect Dis 21: 279–286.

69. Pedrosa J, Saunders BM, Appelberg R, Orme IM, Silva MT, et al. (2000)

Neutrophils play a protective nonphagocytic role in systemic Mycobacterium tuberculosis infection of mice. Infect Immun 68: 577–583.

70. Archer B, Cohen C, Naidoo D, Thomas J, Makunga C, et al. (2009) Interim report on pandemic H1N1 influenza virus infections in South Africa, April to October 2009: epidemiology and factors associated with fatal cases. Euro Surveill 14.

71. Noymer A (2008) The 1918-19 influenza pandemic affected tuberculosis in the United States: reconsidering Bradshaw, Smith, and Blanchard. Biodemography Soc Biol 54: 125–133; discussion 134–140.

72. Housworth J, Langmuir AD (1974) Excess mortality from epidemic influenza, 1957-1966. Am J Epidemiol 100: 40–48.

73. Lofgren S, Callans A (1959) Asian influenza and pulmonary tuberculosis. Acta Med Scand 164: 523–527.

74. Walaza S, Tempia S, Dawood H, Variava E, Moyes J, et al. (2015) Influenza virus infection is associated with increased risk of death amongst patients hospitalized with confirmed pulmonary tuberculosis in South Africa, 2010¿2011. BMC Infect Dis 15: 26.

75. de Paus RA, van Crevel R, van Beek R, Sahiratmadja E, Alisjahbana B, et al. (2013) The influence of influenza virus infections on the development of tuberculosis. Tuberculosis (Edinb) 93: 338–342.

76. Jansen RR, Wieringa J, Koekkoek SM, Visser CE, Pajkrt D, et al. (2011) Frequent detection of respiratory viruses without symptoms: toward defining clinically relevant cutoff values. J Clin Microbiol 49: 2631–2636.

77. Mosser AG, Vrtis R, Burchell L, Lee W-M, Dick CR, et al. (2005) Quantitative and qualitative analysis of rhinovirus infection in bronchial tissues. Am J Respir Crit Care Med 171: 645–651.

78. Venter M, Lassaunière R, Kresfelder TL, Westerberg Y, Visser A (2011) Contribution of common and recently described respiratory viruses to annual hospitalizations in children in South Africa. J Med Virol 83: 1458–1468.

79. Gerna G, Piralla A, Rovida F, Rognoni V, Marchi A, et al. (2009) Correlation of rhinovirus load in the respiratory tract and clinical symptoms in hospitalized immunocompetent and immunocompromised patients. J Med Virol 81: 1498–1507.

80. Freymuth F, Vabret A, Brouard J, Toutain F, Verdon R, et al. (1999) Detection of viral, Chlamydia pneumoniae and Mycoplasma pneumoniae infections in exacerbations of asthma in children. J Clin Virol 13: 131–139.

81. Pientong C, Ekalaksananan T, Teeratakulpisarn J, Tanuwattanachai S, Kongyingyoes B, et al. (2011) Atypical bacterial pathogen infection in children with acute bronchiolitis in northeast Thailand. J Microbiol Immunol Infect 44: 95–100.

82. Ekalaksananan T, Pientong C, Kongyingyoes B, Pairojkul S, Teeratakulpisarn J, et al. (2001) Etiology of acute lower respiratory tract infection in children at Srinagarind Hospital, Khon Kaen, Thailand. Southeast Asian J Trop Med Public Health 32: 513–519.

83. García Ramos E, Pizarro Suárez E, Sapiáin LA, Lugo de la Fuente G [Epidemiologic and etiologic study of acute respiratory infections in children under 5 years of age]. Rev Latinoam Microbiol 33: 109–119.

84. Saikku P, Ruutu P, Leinonen M, Kleemola M, Paladin F, et al. (1993) Mycoplasma pneumoniae and Chlamydia trachomatis in acute lower respiratory infections in Filipino children. Am J Trop Med Hyg 49: 88–92.

85. Principi N, Esposito S, Blasi F, Allegra L (2001) Role of Mycoplasma pneumoniae and Chlamydia pneumoniae in children with community-acquired lower respiratory tract infections. Clin Infect Dis 32: 1281–1289.

86. Muloiwa R, Dube FS, Nicol MP, Hussey G, Zar HJ (2013) 8th World Congress of the World Society for Pediatric Infectious Diseases, Cape Town, South Africa CApe Town.

87. Rendi-Wagner P, Tobias J, Moerman L, Goren S, Bassal R, et al. (2010) The seroepidemiology of Bordetella pertussis in Israel--Estimate of incidence of infection. Vaccine 28: 3285–3290.

88. de Greeff SC, de Melker HE, van Gageldonk PGM, Schellekens JFP, van der Klis FRM, et al. (2010) Seroprevalence of pertussis in The Netherlands: evidence for increased circulation of Bordetella pertussis. PLoS One 5: e14183.

89. Ward JI, Cherry JD, Chang S-J, Partridge S, Keitel W, et al. (2006) Bordetella Pertussis infections in vaccinated and unvaccinated adolescents and adults, as assessed in a national prospective randomized Acellular Pertussis Vaccine Trial (APERT). Clin Infect Dis 43: 151–157.

90. Mooi FR, Van Der Maas NAT, De Melker HE (2014) Pertussis resurgence: waning immunity and pathogen adaptation - two sides of the same coin. Epidemiol Infect 142: 685–694.

91. Hamano-Hasegawa K, Morozumi M, Nakayama E, Chiba N, Murayama SY, et al. (2008) Comprehensive detection of causative pathogens using real-time PCR to diagnose pediatric community-acquired pneumonia. J Infect Chemother 14: 424–432.

92. Simusika P, Bateman AC, Theo A, Kwenda G, Mfula C, et al. (2015) Identification of viral and bacterial pathogens from hospitalized children with severe acute respiratory illness in Lusaka, Zambia, 2011-2012: a cross-sectional study. BMC Infect Dis 15: 52.

93. Cohen C, Walaza S, Moyes J, Groome M, Tempia S, et al. (2014) Epidemiology of Viral-Associated Acute Lower Respiratory Tract Infection among Children <5 Years of Age in a High HIV Prevalence Setting, South Africa, 2009-2012. Pediatr Infect Dis J..

94. Valley-Omar Z, Muloiwa R, Hu N-C, Eley B, Hsiao N-Y (2013) Novel respiratory syncytial virus subtype ON1 among children, Cape Town, South Africa, 2012. Emerg Infect Dis 19: 668–670.

95. Carbonetti NH (2010) Pertussis toxin and adenylate cyclase toxin: key virulence factors of Bordetella pertussis and cell biology tools. Future Microbiol 5: 455–469. Available:

96. Singleton RJ, Bulkow LR, Miernyk K, DeByle C, Pruitt L, et al. (2010) Viral respiratory infections in hospitalized and community control children in Alaska. J Med Virol 82: 1282–1290.

CHAPTER 7

General Conclusions and future prospects

Chapter 1 of this thesis provides an overview of the work presented in this thesis including the rationale.

The work in **Chapter 2** evaluated the recovery of pneumococcus from different types of nasopharyngeal swabs. We showed that pneumococcal recovery from nylon flocked swabs was greater compared to both PET and rayon swabs when testing mock specimens. Similarly, higher

bacterial loads were detected by qPCR from flocked swabs compared with PET swabs from healthy children. It has been shown that the DNA extraction methods may influence the DNA recovery from both cotton and flocked swabs used for the collection of DNA from saliva stains. In this study, we only employed one platform for nucleic acid extraction. Looking ahead, there is a need to assess whether the observed higher bacterial load from flocked swabs was inherent to the specific DNA extraction method used, by performing studies using different nucleic acid extraction methods.

Public health programs use serotype data from invasive pneumococcal disease and carriage to advise vaccine formulation strategies. One of the challenges for pneumococcal surveillance programs is the lack of an inexpensive and robust pneumococcal serotyping method. Assessment of serotyping methods needs to take into account the number of pneumococci to be serotyped, serotype coverage, sample type to be used, how long it takes to perform the test and resource availability. **Chapter 3** therefore described a comparison of a real-time multiplex PCR (rmPCR) to a sequetyping method in order to assess their suitability for high throughput

pneumococcal capsular typing. Whole genome re-sequencing was used to resolve discordant molecular serotyping and reference-standard Quellung results. Although the rmPCR assay performed well for the 21 serotypes/-groups included in the assay, a large proportions of the serotypes encountered in our settings could not be detected. In contrast, the sequetyping assay performed well, with specific exceptions, and may be more useful in settings where vaccine serotypes are less common. We detected structural variation in the *wzh* gene of some strains that were serotyped as 16F by Quellung, but incorrectly sequetyped as 9V by employing whole genome re-sequencing. This highlights a limitation of this approach.

Future work planned is the application of genomic analysis to understand nasopharyngeal colonization, serotype distribution and impact of vaccination during the first year of life. This would specifically address the following objectives:

1) To apply a whole genome sequencing (WGS) approach in understanding the evolution of pneumococcal isolates carried in the nasopharynx over prolonged periods of time at the individual and community level.
2) To use WGS to determine multilocus sequence type in order to define the population structure of pneumococci in this intensively sampled population.
3) To understand the relationship between the presence of bacteriocin cassettes and colonization with other commensal bacteria within the nasopharyngeal microbiome.

Chapter 4 and 5 describe the application of the optimised protocols from chapters 2 and 3 to describe the longitudinal pneumococcal colonisation patterns and serotype distribution during the first year of life among South African infants participating in a birth cohort (Drakenstein Child Health Study). In addition, Chapter 4 details serotype-

specific acquisition rates and duration of pneumococcal carriage. Chapter 5 specifically investigates risk factors for acquisition of pneumococcal carriage including the short-term impact of sequential PCV-13 immunisation on carriage. Given that the "unmeasured" contribution of pneumococcus to non-bacteraemic pneumonia can only be inferred through pneumococcal vaccine studies, the residual burden of pneumococcus to non-bacteraemic pneumonia post PCV implementation remains unclear. We have recently shown no difference in the proportion of children with detectable *S. pneumoniae* in the nasopharynx when comparing children enrolled in a birth cohort who developed pneumonia and controls who did not. However, colonization is likely to mask the true contribution of *S. pneumoniae* to childhood pneumonia. We therefor hypothesize that, since pneumococcal serotypes differ in their invasive potential, if *S. pneumoniae* contributes to the aetiology of pneumonia in this cohort this would manifest as a difference in the distribution of *S. pneumoniae* serotypes between cases and controls. Future work will therefore be to investigate in detail the serotype-specific association with the development of pneumonia in this cohort.

Chapter 6 describes the detection of multiple respiratory microbes in the nasopharynx of children hospitalised with suspected PTB. Whilst there was no clear separation between the pathogens present in the airways of children with and without PTB, *C. pneumoniae*, hMPV, coronavirus O43, influenza C virus, rhinovirus and cytomegalovirus clustered with PTB but this failed to reach statistical significance on testing of each individual microbe. On the other hand, *P. jirovecii, H. influenzae spp*, RSV, *M. pneumoniae*, influenza B virus and enteroviruses clustered towards children without TB although not statistically significant. This pilot work may signal broader differences in the microbial ecology of the upper respiratory tract of these children, which warrants further study. We hypothesize that dysbiosis between regular residents of the upper respiratory tract

(URT) microbiome, that is imbalance between commensals and potential pathogens, is involved in pathogen overgrowth and consequently disease. Further microbiome studies are planned to test this hypothesis.

APPENDICES

APPENDIX 1: Chapter 2 supplementary tables and figures

1.1 CFU validations

Table S1.1: Colony forming units of *Streptococcus pneumoniae* calibrated bacterial cell suspension at mid-log phase

Strain	Dil	Colony forming Units			AVG CFU	Tot DIL Factor	CFU/ml	Average (CFU/ml)
19F	10^{-4}	tmtc	45	49	47	$2.5 * 10^5$	$1.1*10^8$	$1.2*10^8$
	10^{-5}	8	4	3	5	$2.5 * 10^6$	$1.3*10^8$	
	10^{-6}	NG	NG	NG	NA	$2.5 * 10^7$	NA	
1	10^{-4}	56	67	40	54	$2.5 * 10^5$	$1.3*10^8$	$1.2*10^8$
	10^{-5}	16	2	6	8	$2.5 * 10^6$	$2.0*10^8$	
	10^{-6}	NG	1	NG	1	$2.5 * 10^7$	$2.5*10^7$	
5	10^{-4}	29	31	22	27	$2.5 * 10^5$	$6.6*10^7$	$7.1*10^7$
	10^{-5}	2	3	3	3	$2.5 * 10^6$	$7.5*10^7$	
	10^{-6}	NG	NG	NG	NA	$2.5 * 10^7$	NA	

NB: tmtc = too many to count, NG=No growth

Table S1.2: Colony forming units of *Streptococcus pneumoniae* from replicate experiments

Strain	Swabtype	CFU count			Dil.Factor	CFU/ml		
		Day1	Day2	Day3		Day1	Day2	Day3
19F	flocked	71	75	85	$2.5 * 10^6$	$1.78 * 10^8$	$1.88 * 10^8$	$2.13 * 10^8$
		62	67	73	$2.5 * 10^6$	$1.55 * 10^8$	$1.68 * 10^8$	$1.83 * 10^8$
		68	71	79	$2.5 * 10^6$	$1.70 * 10^8$	$1.78 * 10^8$	$1.98 * 10^8$
	Average	67	71	79		$1.68 * 10^8$	$1.78 * 10^8$	$1.98 * 10^8$
	PET	30	32	30	$2.5 * 10^6$	$7.50 * 10^7$	$8.00 * 10^7$	$7.50 * 10^7$
		25	36	34	$2.5 * 10^6$	$6.25 * 10^7$	$9.00 * 10^7$	$8.50 * 10^7$
		22	29	26	$2.5 * 10^6$	$5.50 * 10^7$	$7.25 * 10^7$	$6.50 * 10^7$
	Average	26	32	30		$6.42 * 10^7$	$8.08 * 10^7$	$7.50 * 10^7$
	rayon	9	3	4	$2.5 * 10^6$	$2.25 * 10^7$	$7.50 * 10^6$	$1.00 * 10^7$
		11	4	2	$2.5 * 10^6$	$2.75 * 10^7$	$1.00 * 10^7$	$5.00 * 10^7$
		7	2	4	$2.5 * 10^6$	$1.75 * 10^7$	$5.00 * 10^6$	$1.00 * 10^7$
	Average	9	3	3		$2.25 * 10^7$	$7.50 * 10^7$	$8.33 * 10^7$
	"Control"	68	67	77	$2.5 * 10^6$	$1.70 * 10^8$	$1.68 * 10^8$	$1.93 * 10^8$
		66	71	70	$2.5 * 10^6$	$1.65 * 10^8$	$1.78 * 10^8$	$1.75 * 10^8$
		72	75	82	$2.5 * 10^6$	$1.80 * 10^8$	$1.88 * 10^8$	$2.05 * 10^8$
	Average	69	71	76		$1.72 * 10^8$	$1.78 * 10^8$	$1.91 * 10^8$
1	flocked	82	79	75	$2.5 * 10^6$	$2.05 * 10^8$	$1.98 * 10^8$	$1.88 * 10^8$
		76	73	67	$2.5 * 10^6$	$1.90 * 10^8$	$1.83 * 10^8$	$1.68 * 10^8$
		86	85	71	$2.5 * 10^6$	$2.15 * 10^8$	$2.13 * 10^8$	$1.78 * 10^8$
	Average	81	79	71		$2.03 * 10^8$	$1.98 * 10^8$	$1.78 * 10^8$

	PET	29	29	38	$2.5 * 10^6$	$7.25 * 10^7$	$7.25 * 10^7$	$9.50 * 10^7$
		31	34	42	$2.5 * 10^6$	$7.75 * 10^7$	$8.50 * 10^7$	$1.05 * 10^8$
		30	36	36	$2.5 * 10^6$	$7.50 * 10^7$	$9.00 * 10^7$	$9.00E * 10^7$
	Average	30	33	39		$7.50 * 10^7$	$8.25 * 10^7$	$9.67 * 10^7$
	rayon	6	4	1	$2.5 * 10^6$	$1.50 * 10^7$	$1.00 * 10^7$	$2.50 * 10^6$
		7	7	1	$2.5 * 10^6$	$1.75 * 10^7$	$1.75 * 10^7$	$2.50 * 10^6$
		4	4	1	$2.5 * 10^6$	$1.00 * 10^7$	$1.00 * 10^7$	$2.50 * 10^6$
	Average	6	5	1		$1.42 * 10^7$	$1.25 * 10^7$	$2.50 * 10^7$
	"Control"	77	67	85	$2.5 * 10^6$	$1.93 * 10^8$	$1.68 * 10^8$	$2.13 * 10^8$
		72	71	78	$2.5 * 10^6$	$1.80 * 10^8$	$1.78 * 10^8$	$1.95 * 10^8$
		82	75	81	$2.5 * 10^6$	$2.05 * 10^8$	$1.88 * 10^8$	$2.03 * 10^8$
	Average	77	71	81		$1.93 * 10^8$	$1.78 * 10^8$	$2.03 * 10^8$
5	flocked	20	24	30	$2.5 * 10^6$	$5.00 * 10^7$	$6.00 * 10^7$	$7.50 * 10^7$
		26	18	27	$2.5 * 10^6$	$6.50 * 10^7$	$4.50 * 10^7$	$6.75 * 10^7$
		23	20	22	$2.5 * 10^6$	$5.75 * 10^7$	$5.00 * 10^7$	$5.50 * 10^7$
	Average	23	21	26		$5.75 * 10^7$	$5.17 * 10^7$	$6.58 * 10^7$
	PET	13	3	6	$2.5 * 10^6$	$3.25 * 10^7$	$7.50 * 10^6$	$1.50 * 10^7$
		19	1	7	$2.5 * 10^6$	$4.75 * 10^7$	$2,50 * 10^6$	1.75
		17	2	5	$2.5 * 10^6$	$4.25 * 10^7$	$5.00 * 10^6$	$1.25 * 10^7$
	Average	16	2	6		$4.08 * 10^7$	$5.00 * 10^6$	$1.50 * 10^7$
	rayon	3	4	3	$2.5 * 10^6$	$7.50 * 10^6$	$1.00 * 10^7$	$7.50 * 10^6$
		1	2	5	$2.5 * 10^6$	$2.50 * 10^6$	$5.00 * 10^6$	$1.25 * 10^7$
		2	3	1	$2.5 * 10^6$	$5.00 * 10^6$	$7.50 * 10^6$	$2.50 * 10^6$
	Average	2	3	3		$5.00 * 10^6$	$7.50 * 10^6$	$7.50 * 10^6$
	"Control"	19	26	29	$2.5 * 10^6$	$4.75 * 10^7$	$6.50 * 10^7$	$7.25 * 10^7$
		11	25	33	$2.5 * 10^6$	$2.75 * 10^7$	$6.25 * 10^7$	8.25
		16	22	30	$2.5 * 10^6$	$4,00 * 10^7$	$5,50 * 10^7$	$7,50 * 10^7$
	Average	15	24	31		$3.83 * 10^7$	$6.08 * 10^7$	$7.67 * 10^7$

Table S1.3: NP swabs comparisons - *Streptococcus pneumoniae* ATCC 49619 (serotype 19F)

Type of NP Swab	Dil	Colony forming Units			AVG CFU	Tot DIL Factor	CFU/ml
Flocked	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^5$	NA
	10^{-3}	67	71	79	71	$2.5 * 10^6$	$1.8*10^8$
	10^{-4}	4	11	6	7	$2.5 * 10^7$	$1.8*10^8$
Dacron	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^5$	NA
	10^{-3}	26	32	30	29	$2.5 * 10^6$	$7.3*10^7$
	10^{-4}	1	1	1	1	$2.5 * 10^7$	$2.5*10^7$
Rayon	10^{-2}	78	44	86	69	$2.5 * 10^5$	$1.7*10^7$
	10^{-3}	9	3	3	5	$2.5 * 10^6$	$1.3*10^7$
	10^{-4}	NG	NG	NG	NA	$2.5 * 10^7$	NA
"Control"	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^5$	NA
	10^{-3}	69	71	76	72	$2.5 * 10^6$	$1.8*10^8$
	10^{-4}	6	11	11	9	$2.5 * 10^7$	$2.3*10^8$

NB: tmtc = too many to count, NG=No growth

Table S1.4: NP swabs comparisons - *Streptococcus pneumoniae serotype 1*

Type of NP Swab	Dil	Colony forming Units			AVG CFU	Tot DIL Factor	CFU/ml
Flocked	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^{5}$	NA
	10^{-3}	tmtc	tmtc	tmtc	NA	$2.5 * 10^{6}$	NA
	10^{-4}	81	79	70	77	$2.5 * 10^{7}$	$1.9*10^{9}$
Dacron	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^{5}$	NA
	10^{-3}	116	121	128	122	$2.5 * 10^{6}$	$3.1*10^{8}$
	10^{-4}	30	33	39	34	$2.5 * 10^{7}$	$8.5*10^{8}$
Rayon	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^{5}$	NA
	10^{-3}	59	26	34	40	$2.5 * 10^{6}$	$1.0*10^{8}$
	10^{-4}	6	5	1	4	$2.5 * 10^{7}$	$1.0*10^{8}$
"Control"	10^{-2}	tmtc	tmtc	tmtc	NA	$2.5 * 10^{5}$	NA
	10^{-3}	tmtc	tmtc	tmtc	NA	$2.5 * 10^{6}$	NA
	10^{-4}	77	71	83	77	$2.5 * 10^{7}$	$1.9*10^{9}$

NB: tmtc = too many to count, NG=No growth

Table S1.5: NP swabs comparisons - *Streptococcus pneumoniae serotype 5*

Type of NP Swab	Dil	Colony forming Units			AVG CFU	Tot DIL Factor	CFU/ml
Flocked	10^{-2}	tmtc	116	105	111	$2.5 * 10^{5}$	$2.8*10^{7}$
	10^{-3}	23	21	26	23	$2.5 * 10^{6}$	$5.8*10^{7}$
	10^{-4}	3	2	4	3	$2.5 * 10^{7}$	$7.5*10^{7}$
Dacron	10^{-2}	56	67	40	54	$2.5 * 10^{5}$	$1.3*10^{7}$
	10^{-3}	16	2	6	8	$2.5 * 10^{6}$	$2.0*10^{7}$
	10^{-4}	NG	1	NG	1	$2.5 * 10^{7}$	$2.5*10^{7}$
Rayon	10^{-2}	29	31	22	27	$2.5 * 10^{5}$	$6.6*10^{6}$
	10^{-3}	2	3	3	3	$2.5 * 10^{6}$	$7.5*10^{6}$
	10^{-4}	NG	NG	NG	NA	$2.5 * 10^{7}$	NA
"Control"	10^{-2}	83	89	80	84	$2.5 * 10^{5}$	$2.1*10^{7}$
	10^{-3}	22	23	25	23	$2.5 * 10^{6}$	$5.8*10^{7}$
	10^{-4}	6	2	2	3	$2.5 * 10^{7}$	$7.5*10^{7}$

NB: tmtc = too many to count, NG=No growth

Table S1.6: Recovery of *S. pneumoniae* from mock specimens using flocked, PET (Dacron™) and rayon swabs

S. pneumoniae strain ("Control")[a]	Swab type	n[b]	Mdian CFU recovery/ml (IQR)[c]	Percentage recovery [d]
Serotype 19F[e] (18×10^7)	Flocked	9	17.8×10^7 ($17 \times 10^7 - 18.8 \times 10^7$)	99 (94-105)
	PET	9	7.8×10^7 (6.5×10^7 - 8×10^7)	43 (36-44)
	rayon	9	1.0×10^7 ($0.8 \times 10^7 - 1.8 \times 10^7$)	6 (4-10)
Serotype 1 (19×10^7)	Flocked	9	19×10^7 ($18.3 \times 10^7 - 20.5 \times 10^7$)	100 (96-107)
	PET	9	8.5×10^7 (7.5×10^7 - 9×10^7)	45 (39-47)
	Rayon	9	1.0×10^7 ($0.25 \times 10^7 - 1.5 \times 10^7$)	5 (1-8)
Serotype 5 (5.8×10^7)	Flocked	9	5.8×10^7 ($5 \times 10^7 - 6.5 \times 10^7$)	100 (86-112)
	PET	9	1.5×10^7 ($0.8 \times 10^7 - 3.3 \times 10^7$)	26 (14-57)
	Rayon	9	0.8×10^7 ($0.5 \times 10^7 - 0.8 \times 10^7$)	14 (9-14)

SD, Standard deviation

[a] Represents a simulated 100% release of *S. pneumoniae* into skim milk-tryptone-glucose-glycerol (STGG) media and is used as a reference to calculate the percentage of recovery. [b] Pooled replicates of each swab type from three independent experiments performed over three different days. [c]Median of colony forming units (CFU) obtained from 9 different replicates of each swab type tested, IQR=interquartile range. [d]Percentage of recovery was calculated as the proportion of the median CFU recovered from each swab type divided by the control (100% CFU recovery) for each respective *S. pneumoniae* serotype. [e]*S. pneumoniae* serotype 19F tested in this study corresponds to the American Type Culture collection (ATCC 49619) strain (21).

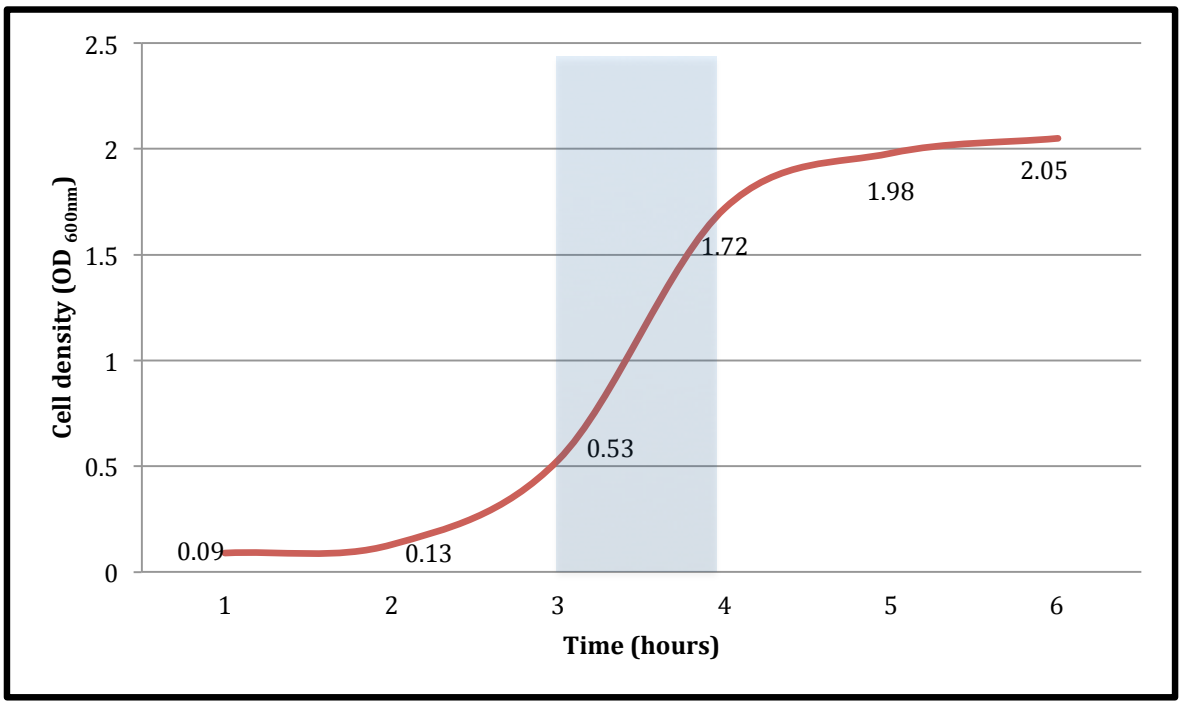

Figure S1.1: *Streptococcus pneumoniae* ATCC 49619 growth curve plotting the cell density

(OD_{600nm} measurement) against time. Growth curve was used to determine the mid-log exponential) growth phase of S. *pneumoniae* ATCC 49619. The exponential phase was determined to be between 3 and 4 hours, as indicated by the light blue solid bar on the graph.

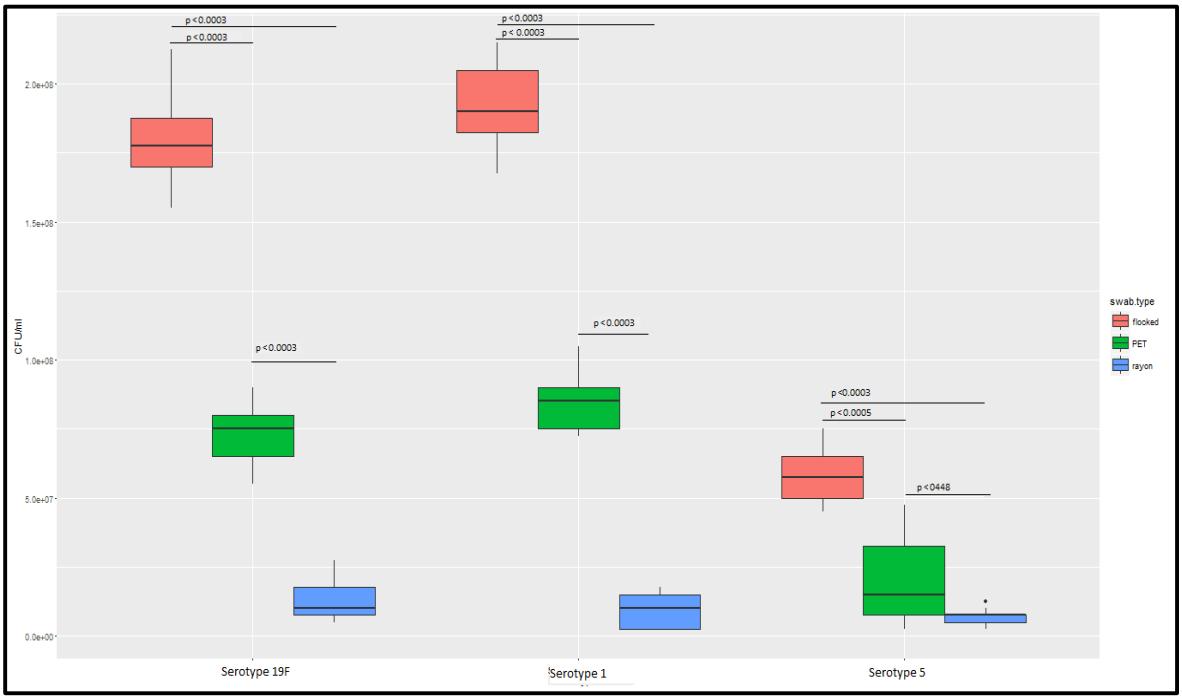

Figure S1.2: Box and whisker plots for the recovery of *Streptococcus pneumoniae* from mock specimens using nylon flocked, PET (Dacron™) and rayon swabs. Red=Flocked swabs, Green=PET, Blue=rayon swabs. The data is presented as the pooled median CFU recovered from 9 replicates of each swab type using (Tables S1.3-2.5). Mann-whitney rank test was used to compare median colony forming units/ml (CFU/ml) recovered from each swab types. p-values adjusted for multiple comparisons.

APPENDIX 2: Chapter 3 supplementary tables and figures

2.1 rmPCR Assay

Table S2.1: Primer and probe data for the sequential real-time multiplex PCR assay

GenBank acc. no.(*gene*)	Primer/probe ID	Primer/probe sequence (5' -3')	Probe dye	Probe special chemistry	Quencher (3')	nM
CR931632 (*wzy*)	1-F	TTTCATCCCTATGTGTGGTATAG	FAM	LNA[b]	BHQ1	300
	1-R	GCTTTAGAAGGTAGAGTTAACAAC				300
	1-**Probe**	TGC**C**AA**A**GC**C**AG**C**CAT				100
CR931633 (*wzy*)	2-F	TGTTATCCCATATAAGAACCGAGTGT	FAM	"T"[d=]*BHQ1*	BHQ1	300
	2-R	AAAATTACCCCAAAAGCTATCCAA				300
	2-**Probe**	TTGCAATT"T"CAATTTTTTTGCCCCAATCTC				200
CR931634 (*galU*)	3-F	CCACTAAAGCTTTGGCAAAAGAAA	HEX	"T"[d=]*BHQ1*	BHQ1	300
	3-R	CCCGAACGTAAAGCTTCTTCA				300
	3-**Probe**	TTGTAGACCGCCCCACAA"T"TCATTTTGT				200
CR931635 (*wzy*)	4-F	GCTTCTGCTGTAACTGTTGTGC	CY5		BHQ2	300
	4-R	CACCACCATAGTAACCAAAGTTCC				300
	4-**Probe**	TTCCACAAAAGAAGAGCCTACAGGTAACCCCA				100
CR931637 (wzy)	5-F	CATGATTTATGCCCTCTTGCAA	HEX	"T"[d=]*BHQ1*	BHQ1	300
	5-R	GACAGTATAAGAAAAAGCAAGGGCTAA				300
	5-**Probe**	TCTTCTTCTCA"T"CGTTTCCGCATGCTTTT				200
CR931639 (*wciP*)	6A/6B/6C/6D-F	GTTTGCACTAGAGTATGGGAAGG	FAM	"T"[d=]*BHQ1*	BHQ1	200
	6A/6B/6C/6D-R	TAGCCTTTCTGAAAACATTTAGCG				200
	6A/6B/6C/6D-**Probe**	TGTTCTGCCC"T"GAGCAACTGGTCTTGTATC				200
EF538714 (*wciN*)	6C/6D-F	TTGGGATGATTGGTCGTATTAG	FAM	LNA[b]	BHQ1	200
	6C/6D-R	CTCTTCAATTAGTTCTTCAGTTCG				200
	6C/6D-**Probe**	CCA**C**GC**A**AT**T**CG**C**CATC				100
CR931643 (*wzy*)	7F/7A-F	ATGAAGGCTTTGGTTTGACAGG	CY5		BHQ2	200
	7F/7A-R	ATTCTCGCCATCAATTGCATATTC				200
	7F/7A-**Probe**	ACACCACTATAGGCTGTTGAGACTAACGCACA				100
CR931648 (*wzx*)	9V/9A-F	AGGTATCCTATATACTGCTTTAGG	HEX	LNA[b]	BHQ1	300
	9V/9A-R	CGAATCTGCCAATATCTGAAAG				300
	9V/9A-**Probe**	ACA**C**AT**T**GA**C**AA**C**CGCT				100
CR931653 (*wzy*)	11A/11D-F	AAATGGTTTGGATATGGTTTGTTTGG	CY5		BHQ2	300
	11A/11D-R	AGTGCTAACTGTAAAACTTGATTATGAG				300
	11A/11D-**Probe**	ATTCCAACTTCTCCCAATTTCTGCCACGG				100
CR931660 (*wzx*)	12F/12A/12B/44/46-F	GCACCCACGGGTAAATATTCTAC	CY5		BHQ2	300
	12F/12A/12B/44/46-R	CAACTAAGAACCAAGGATCCACAG				300
	12F/12A/12B/44/46-**Probe**	TGCCCACCAACACCAGGTCCAGGT				200
CR931662 (*wzy*)	14-F	AGAGTGTATGAGGAATCC	FAM	"T"[d=]*BHQ1*	BHQ1	300
	14-R	ATATATCTACTGTAGAGGGAAT				300

	14-**Probe**	CGCCAAGTAACA“T”TTCCATTCCATT				100
CR931663 (wzy)	15A/15F-F	AATTGCCTATAAACTCATTGAGATAG				200
	15A/15F-R	CCATAGGAAGGAAATAGTATTTGTTC	FAM	LNA[b]	BHQ1	200
	15A/15F-**Probe**	CCC**G**CA**A**AC**T**CT**G**TCCT				100
CR931668 (*wzy*)	16F-F	TAATGTTATGACCTTGGTAATCTTCCC				300
	16F-R	TCCCAAAGGATAATCAATAACTTTTAGAAG				300
	16F-**Probe**	AGCCATAAGTCT“T”CCAAATGCTTAACCGCT	HEX	“T”[d=]*BHQ1*	BHQ1	100
CR931673 (*wzy*)	18C/18A/18B//18F-F	TCGATGGCTAGAACAGATTTATGG				200
	18C/18A/18B//18F-R	CCATTGTCCCTGTAAGACCATTG	HEX		BHQ1	200
	18C/18A/18B//18F-**Probe**	AGGGAGTTGAATCAACCTATAATTTCGCCCC				100
CR931675 (*wzy*)	19A-F	CGCCTAGTCTAAATACCA				200
	19A-R	GAGGTCAACTATAATAGTAAGAG	FAM		BHQ1	200
	19A-**Probe**	TATCAATGAGCCGATCCGTCACTT				100
CR931678 (*wzy*)	19F-F	TGAGGTTAAGATTGCTGATCG				300
	19F-R	CACGAATGAGAACTCGAATAAAAG	CY5	LNA[b]	BHQ1	300
	19F-**Probe**	CGC**A**CT**G**TC**A**AT**T**CACCTTC				100
CR931682 (*wcwV*)	22F/22A-F	TCTATTAAATAACCCATTGGAATTGAAACG				200
	22F/22A-R	TCGCAATTGAAGACCACATAAACTG	HEX	“T”[d=]*BHQ1*	BHQ1	200
	22F/22A-**Probe**	TCCGTAAT“T”CGCTTATGGGCACATTCTCCA				200
CR931683 (*wzy*)	23A-F	CTCCCCTCCATTACCCATTTGG	CY5	“T”[d=]*BHQ2*	BHQ2	200
	23A-R	TGAAGAAAGTGCTGTTTGTGAACC				200
	23A-**Probe**	AGCTAGAAC“T”CCCACACTCCCTACTCCCA				100
CR931685 (*wzy*)	23F-F	GACAGCAACGACAATAGTCATCTC				300
	23F-R	TCCATCCCAACCTAACACACTTC	CY5	“T”[d=]*BHQ2*	BHQ2	300
	23F-**Probe**	ATTGTGTCCA“T”AACCCTTCGTCGTATTTCCAAAG				200
CR931702 (*wzy*)	33F/33A/37-F	GGAACTGGTTCAGCAACTATACG				200
	33F/33A/37-R	GGTTCTAAGACCGTCTGAAATACC	HEX	“T”[d=]*BHQ1*	BHQ1	200
	33F/33A/37-**Probe**	CCCCAAATAGGAC“T”TTTCTGCCATGCCAAA				200

[a]Abbreviations: FAM, 6-carboxyfluorescein; HEX, hexachloro-6-
cein; ROX, 6-carboxy-X-rhodamine; CY5, indodicarbocyanine; BHQ, black hole
dentification; LNA, locked nucleic acid. [b]Locked nucleic acid nucleotides are
Γ” black hole quencher placed internally on the thymidine, highlighted with a green
able adopted from Pimenta *et al* (32).

Table S2.2: Reaction 1 (*Pneumococcal* serotypes **1*, 5** and **23F**) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
2x Mastermix	12,5	500	1X
1-F (10uM)	0,75	30	300nM
1-R (10uM)	0,75	30	300nM
1-pb FAM (10uM)	0,25	10	100nM
5-F (10uM)	0,75	30	300nM
5-R (10uM)	0,75	30	300nM
5-pb HEX (10uM)	0,50	20	200nM
23F-F (10uM)	0,75	30	200nM

23F-R (10uM)	0,75	30	200nM
23F-pb Cy5 (10uM)	0,5	20	100nM
H2O	1,75	70	
DNA	5		
Total	25	800	

Table S2.3: Reaction 2 (*Pneumococcal* serotypes **4***, **6A/6B/6C/6D, 9V/9A**) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
4-F (10 uM)	0,75	30	300nM
4-R (10 uM)	0,75	30	300nM
4-pb (10 uM)	0,25	10	100nM
6A/6B/6C/6D-F (10 uM)	0,5	20	200nM
6A/6B/6C/6D-R (10 uM)	0,5	20	200nM
6A/6B/6C/6D-pb (10 uM)	0,5	20	200nM
9V/9A-F (10 uM)	0,75	30	300nM
9V/9A-R (10 uM)	0,75	30	300nM
9V/9A-pb (10 uM)	0,25	10	100nM
Mastermix	12,5	500	1X
H2O	2,5	100	
DNA	5		
Total	25	800	

Table S2.4: Reaction 3 (*Pneumococcal* serotypes **14*, 18A/18B/18C/18F, 19F**) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
14-F (10 uM)	0,75	30	300nM
14-R (10 uM)	0,75	30	300nM
14-pb (10 uM)	0,25	10	100nM
18C/18F/18B/18A-F (10 uM)	0,5	20	200nM
18C/18F/18B/18A-R (10 uM)	0,5	20	200nM
18C/18F/18B/18A-pb (10 uM)	0,25	10	100nM
19F-F (10 uM)	0,75	30	300nM
19F-R (10 uM)	0,75	30	300nM
19F-pb (10 uM)	0,25	10	100nM
Mastermix	12,5	500	1X
H2O	2,75	110	
DNA	5		
Total	25	800	

Table S2.5: Reaction 4 (*Pneumococcal* serotypes **3*, 7F/7A,19A**) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
3-F (10uM)	0,75	30	300nM
3-R (10uM)	0,75	30	300nM
3-pb HEX (10uM)	0,5	20	200nM
7F/7A-F (10uM)	0,5	20	200nM
7F/7A-R (10uM)	0,5	20	200nM
7F/7A-pb ROX (10uM)	0,25	10	100nM
19A-F (10uM)	0,5	20	200nM
19A-R (10uM)	0,5	20	200nM
19A-pb FAM (10uM)	0,25	10	100nM

Mastermix	12,5	500	1X
H2O	3	120	
DNA	5		
Total	25	800	

Table S2.6: Reaction 5 (*Pneumococcal* serotypes 6C/6D, 12F/12A/12B/44/46, 22F/22A) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
6C/6D-F (10uM)	0,5	20	200nM
6C/6D-R (10uM)	0,5	20	200nM
6C/6D-pb FAM (10uM)	0,25	10	100nM
12F/12A/12B/44/46-F (10uM)	0,75	30	300nM
12F/12A/12B/44/46-R (10uM)	0,75	30	300nM
12F/12A/12B/44/46-pb ROX (10uM)	0,5	20	200nM
22F/22A-F (10uM)	0,5	20	200nM
22F/22A-R (10uM)	0,5	20	200nM
22F/22A-pb HEX (10uM)	0,5	20	200nM
Mastermix	12,5	500	1X
H2O	2,75	110	
DNA	5		
total	25	800	

Table S2.7: Reaction 6 (*Pneumococcal* serotypes 15A/15F, 23A,33F/33A/37) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
15A/15F-F (10 uM)	0,5	20	300nM
15A/15F-R (10 uM)	0,5	20	300nM
15A/15F-pb (10 uM)	0,25	10	200nM
23A-F (10 uM)	0,5	20	200nM
23A-R (10 uM)	0,5	20	200nM
23A-pb (10 uM)	0,25	10	100nM
33F/33A/37-F (10 uM)	0,5	20	200nM
33F/33A/37-R (10 uM)	0,5	20	200nM
33F/33A/37-pb (10 uM)	0,5	20	100nM
Mastermix	12,5	500	1X
H2O	3,5	140	
DNA	5		
Total	25	800	

Table S2.8**:** Reaction 7 (*Pneumococcal* serotypes 2, 11A/11D,16F) for the African region rmPCR mix for 40 reactions

Reagent (initial concentration)	1 Reaction (µl)	X40 Reactions (µl)	Final Concentration
2-F (10uM)	0,75	30	300nM
2-R (10uM)	0,75	30	300nM
2-pb FAM (10uM)	0,5	20	200nM
11A/11D-F (10uM)	0,75	30	200nM
11A/11D-R (10uM)	0,75	30	200nM
11A/11D-pb ROX (10uM)	0,25	10	100nM
16F-F (10uM)	0,75	30	200nM
16F-R (10uM)	0,75	30	200nM

16F-pb HEX (10uM)	0,25	10	100nM
Mastermix	12,5	500	1X
H2O	2	80	
DNA	5		
Total	25	800	

*PCV-13 serotypes

3.2 Sequetyping Assay

Table S2.9: Sequetyping assay PCR reaction mix for 80 reactions

Reagent (initial concentration)	1 Reaction (µl)	X80 Reactions (µl)	Final Concentration
2x Mastermix	12.5	1000	1X
Cps Primer_F 10 uM	1	80	400nM
Cps Primer_R 10 uM	1	80	400nM
dH2O	8.5	680	
DNA	2		
total	25	2000	

Table S2.10: Pneumococcal strain names and their corresponding serotypes.

Strain name	Serotype
SPN033038	1
SPN032672	1
INV104	1
P1031	1
OXC141	3
SPN034156	3
SPN034183	3
SPN994038	3
SPN994039	3
SPNA45	3
TIGR4	4
70585	5
CGSP14	14
INV200	14
JJA	14
6320	20
5931-06	20
030-1341B	10A
AP200	11A
106-498B	15B/15C
G54	19F
ST556	19F

Taiwan 19F-14	19F
PO-329	19F
2236/42	24B
CHPA388	6C
MNZ21	6D
MNZ920	6D

2.3 How to run the Sequetyper program

In order to assign serotypes, download the sequetyper program from http://www.gematics.com/sequetyper.html, which is stored as a zip folder. After extracting the program to a specified location:

1) Navigate to the sequetyper folder and double click the run in order to launch the program for Windows users or use the run. Command for MAC operating system users.

2) Click on “File” and then “Open directory”

3) Now navigate to a specified directory that contains the AB1 files which have been labelled using a set convention (See under rebuilding the database for the naming format).

4) Once you have selected a directory, click on the “Open” button

5) A window will open to notify the user if there are any errors in file processing, which can include sequences that for below the 0.05 PHRED cut-off or sequences that are duplicated or have a missing compliment sequence.

6) If this QC step is resolved, click on “Proceed anyway”

7) The program will then find all the alignments for the input files.

8) Once the alignment is complete, a report window will open showing the results.

Rebuilding the Sequetyper database

In order to rebuild the database of reference sequences the following steps must be carried out:

1) First locate the file called "categoryCatAll.fasta" located in the "db" folder,

2) Open this file for editing in a text editor and

3) Paste in new sequence in fasta format.

Note: The new sequence must have a specific format. The specific format applies to the definition line, the line with the ">" in it. This normally looks something like: *>gi|68643586|emb|CR931683.1|Gene locus A*
It should have the following format:
>w|x|y|z|serotype_A_B_C_2#D
where

- >w|x|y|z|can be anything but
- _A_ is the Pneumococcal serogroup,
- _B _Pneumococcal serotype, which could be designated as F, A, B, C, D, N, L, V
- or left blank for serotypes that do not have any serotypes included within the serogroup,
- _C_is the allele number if applicable and
- _D_ is the Pneumococcal sequence description.

An example of what it should look like is:
>gi|68643586|emb|CR931683.1|serotype_23_a_0_2#Streptococcus pneumoniae strain 1196/45 (serotype 23A) 21,475 bp linear DNA.
The definition line must contain four underscores (_) and one hash (#).
After the sequence has been added, or sequences have been removed, save the file and return to the program. Then

4) Click on 'File' and then 'Rebuild database'

This will rebuild the database using the sequences as they appear in the fasta file. The program will then use the updated information when assigning serotypes.

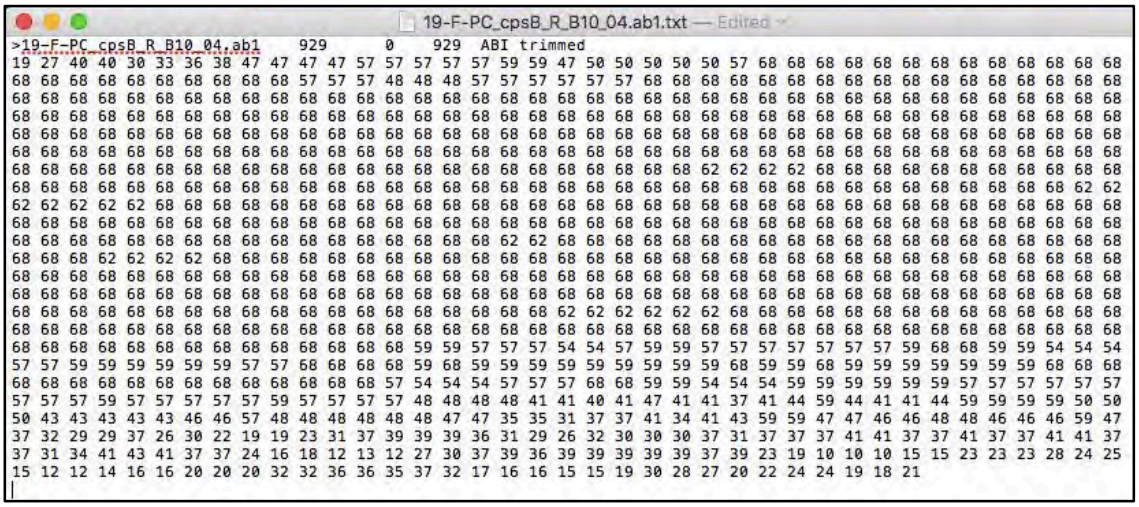

Fig S2.1: Example of Phred score export generated using Sequetyper program

APPENDIX 3: Chapter 4 supplementary tables and figures

3.1 Methodology: Serotype-specific inter-dependency by functional data analysis

We applied functional data analysis to explore the inter-strain- and serotype specific relationship using our longitudinal serotype distribution data. In multiple regression, a single response, y, is predicted by p predictor variables, $x_1, x_2, \dots, x_p$.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \qquad (214)$$

In functional data analysis, observations are in the form of continuous functions. In the case here, $x(t)$ is the presence / absence of serotype x over the continuous time period from birth to one year. Discreet observations are made of $x(t)$ at time points $t = birth, 2\ weeks, 4\ weeks, \dots, 52\ weeks$.

Functional regression analysis with functional response and functional predictors, replace the observations in (1) with continuous functions $y(t)$ for the response and $x_1(t), \dots, x_p(t)$ for the predictors. Here, each serotype is in turn modeled as a response, with all the other serotypes are predictors. A naïve form of the functional regression model will be

$$y(t) = \beta_0(t) + \beta_1(t)x_1(t) + \cdots + \beta_p(t)x_p(t) + \varepsilon(t) \quad (213)$$

In (2), the presence/absence of serotype y at time point t, depends on the presence/absence of serotypes $x_1, x_2, \dots, x_p$ at time point t. A more realistic model, models the presence/absence of serotype y at time point t, as a function of the presence/absence of the predictor serotypes at all time points from birth to time point t. We shall denote the presence/absence of predictor serotype i over this period as $x_i(s), 0 \le s \le t$.

The regression model as given by Ramsay, Hooker and Graves (2009) is now of the form

$$y(t) = \beta_0(t) + \int_0^t \beta_1(s,t)x_1(s)ds + \cdots + \int_0^t \beta_p(s,t)x_p(s)ds + \varepsilon(t) \quad (379)$$

The integral in each predictor term in (3), essentially combines the effect of x_i from birth up to time point t. Mathematically an integral can be interpreted as the area under the integrant, i.e. the area under the function $\beta_i(s,t)x_i(s)$ between the points $s = 0$ and $s = t$. The fitting of the regression function is implemented in the fda package in R (213,379).

As t varies from birth to 52 weeks, $y(t)$ changes over time. The changes in $y(t)$ is now modelled by an intercept term $\beta_0(t)$, which is also a function over $0 < t < 52$ and the predictors. $y(0)$ can only be predicted from the values of the predictors at time $s = 0$. As t increases towards $t = 52$, a larger spread of the predictor functions, $x_i(s), 0 \le s \le t$, could possibly predict $y(t)$. The regression coefficients are therefore depending on

the value of t, and all the values of s, $0 \leq s \leq t$, spreading a triangular region. As in ordinary multiple regression, the size of the regression coefficients indicates the magnitude of the effect of the predictor on the response. A coefficient value of zero, indicates that the predictor has no effect on the response, a large positive value indicates a positive effect, i.e. an increase in the predictor will lead to an increase in the response and negative regression coefficients vice versa. Ramsay, *et al*. (2009) does not provide any inference methodology for testing the statistical significance of coefficients. Also, there is no indication of the meaning of the numerical values of the coefficients, since it should be interpreted in terms of the area under a function between $s = 0$ and $s = t$.

Here coefficients are simply explored graphically with a contour heat map. Shades close to transparent are viewed as almost zero and only the most noticeable coefficients are interpreted.

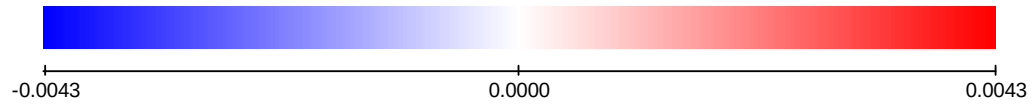

Figure S4.1: Colour grids utilised in exploring the strength of pneumococcal serotype inter-dependency relationships. Transparent zones, shows no relationship, while the red gradient represent increased probability of acquiring one serotype given the other. The converse is true for the blue gradient, that is, the present of one serotype reduces the probability of acquiring the other.

APPENDIX 4: Chapter 5 supplementary tables and figures

4.2 Cox proportion hazard models

Table S4.1: Cox proportional hazards model of time to first carriage acquisition of ANY pneumococcal serotype, stratified by site.

Risk factors		TC Newman (n=59)			Mbekweni (n=77)			Crude (n=136)		
		HR	**95% CI**	***p***	**HR**	**95% CI**	***p***	**HR**	**95% CI**	***p***
Gender	Female	1	-	-	1	-	-	1	-	-
	Male	0.97	0.77-1.22	0.77	1.07	0.88-1.31	0.50	1.03	0.89-1.19	0.720
Mode of delivery	Normal vaginal	1	-	-	1	-	-	1	-	-
	Normal vacuum	**0.30**	**0.25-0.36**	**0.00**	-	-	-	0.30	0.27-0.34	**0.00**
	Elective caesarean	0.63	0.27-1.48	0.29	1.24	0.89-1.71	0.20	1.01	0.70-1.45	0.97
	Emergency caesarean	0.98	0.77-1.26	0.89	0.92	0.79-1.06	0.25	0.96	0.83-1.12	0.61
HIV exposure	No	1	-	-	1	-	-	1	-	-
	Yes	**1.47**	**1.11-1.95**	**0.00**	0.91	0.74-1.12	0.38	0.99	0.82-1.19	0.89
Feed choice	Exclusive breastfed	1	-	-	1	-	-	1	-	-
	Mixed fed	1.19	0.92-1.55	0.19	0.97	0.70-1.33	0.84	1.05	0.85-1.31	0.64
	Never breastfed	0.89	0.69-1.16	0.40	0.90	0.70-1.15	0.40	0.91	0.76-1.08	0.29
Other Children[1]	No	1	-	-	1	-	-	1	-	-
	Yes	**1.49**	**1.21-1.85**	**0.00**	1.07	0.88-1.31	0.51	**1.23**	**1.06-1.43**	**0.00**
Premature	No	1	-	-	1	-	-	1	-	-
	Yes	1.17	0.89-1.55	0.26	0.76	0.54-1.07	0.12	0.94	0.75-1.19	0.60
Smoking status of Mother	No	1	-	-	1	-	-	1	-	-
	Yes	1.22	0.88-1.68	0.23	1.01	0.82-1.24	0.94	1.05	0.90-1.22	0.56
Sleeps with Mother	No	1	-	-	1	-	-	1	-	-
	Yes	1.24	0.67-2.29	0.50	**0.72**	**0.56-0.94**	**0.01**	0.84	0.65-1.09	0.20
Sleeps with Father	No	1	-	-	1	-	-	1	-	-
	Yes	1.17	0.93-1.48	0.19	1.20	0.98-1.47	0.08	1.17	1.01-1.36	**0.03**
Sleeps with sibling	No	1	-	-	1	-	-	1	-	-
	Yes	**1.50**	**1.18-1.89**	**0.00**	1.15	0.95-1.39	0.16	1.28	1.11-1.48	**0.00**
Microwave	No	1	-	-	1	-	-	1	-	-
	Yes	0.98	0.59-1.63	0.94	**0.79**	**0.64-0.98**	**0.03**	0.83	0.69-1.00	**0.05**
Paraffin	No	1	-	-	1	-	-	1	-	-
	Yes	1	-	-	1.12	0.90-1.38	0.30	1.12	0.92-1.35	0.27
Wood	No	1	-	-	1	-	-	1	-	-
	Yes	1.27	0.91-1.77	0.17	1	-	-	1.24	0.92-1.69	0.16
Gas	No	1	-	-	1	-	-	1	-	-
	Yes	0.74	0.47-1.17	0.20	0.62	0.22-1.71	0.35	0.69	0.43-1.10	0.12

*One observation with excluded for missing data. [1]Other children staying with the child in the same household. "-" Insufficient data in that category to make any inferences

In order to clearly determine which pneumococcal serotypes was actually acquired by these children, the CPH models were extended to categorically determine which risk factors were significant drivers of VT or NVT carriage acquisition, table 4. HIV exposure and presence of older siblings were associated with increased risk of time to first episode of VT pneumococcal acquisition amongst children at TC Newman (HR = 4.12, 95% CI = 1.21 - 14.03, p = 0.02 and HR = 2.57, 95% CI = 1.24 - 5.30, p = 0.01 respectively). In contrast, HIV exposure and presence of older siblings were not significant drivers of VT acquisition at Mbekweni, (HR = 0.97, 95% CI = 0.54 - 1.74, p = 0.92 and HR = 1.36, 95% CI = 0.81 - 2.26, p = 0.24 respectively). In Mbekweni, older had to be sleeping in the same room as the child in order for them to be potential sources of VT transmission, HR = 1.66, 95% CI = 1.00 - 2.73, p = 0.04.

Table S4.2: Cox proportional hazards model of time to first carriage acquisition of PCV-13 vaccine serotypes, stratified by site.

Risk factors		**TC Newman (n=59)**			**Mbekweni (n=77)**			**Crude (n=136)**		
		HR	**95% CI**	***p***	**HR**	**95% CI**	***p***	**HR**	**95% CI**	***p***
Gender	Female	1	-	-	1	-	-	1	-	-
	Male	0.83	0.44-1.58	0.58	1.64	0.89-3.01	0.11	1.19	0.677-1.85	0.43
Mode of delivery	Normal vaginal	1	-	-	1	-	-	1	-	-
	Normal vacuum	1.17	0.72-1.90	0.53	1	-	-	1.26	0.89-1.79	0.19
	Elective caesarean	0.95	0.28-3.21	0.94	1.55	0.66-3.62	0.32	1.26	0.63-2.51	0.51
	Emergency caesarean	0.72	0.30-1.74	0.47	0.36	0.06-2.29	0.28	0.60	0.26-1.37	0.23
HIV exposure	No	1	-	-	1	-	-	1	-	-
	Yes	**2.41**	**1.46-3.98**	**0.00**	0.99	0.53-1.85	0.97	1.07	0.65-1.77	0.78
Feed choice	Exclusive breastfed	1	-	-	1	-	-	1	-	-
	Mixed fed	0.76	0.23-2.50	0.65	1.17	0.49-2.81	0.73	0.86	0.45-1.65	0.65
	Never breastfed	**0.48**	**0.26-0.87**	**0.01**	0.99	0.44-2.21	0.98	0.67	0.42-1.07	0.09
Other Children[1]	No	1	-	-	1	-	-	1	-	-
	Yes	**2.67**	**1.48-4.83**	**0.00**	0.72	0.37-1.40	0.34	1.35	0.86-2.11	0.19
Premature	No	1	-	-	1	-	-	1	-	-
	Yes	1.53	0.69-3.39	0.30	1.09	0.56-2.14	0.79	1.32	0.76-2.29	0.33
Smoking status of Mother	No	1	-	-	1	-	-	1	-	-
	Yes	2.46	1.07-5.69	**0.03**	0.58	0.31-1.08	0.08	0.99	0.63-1.55	0.97
Sleeps	No	1	-	-	1	-	-	1	-	-

with Mother	Yes	-	-	-	**0.46**	**0.25-0.85**	**0.01**	0.89	0.46-1.74	0.74
Sleeps with Father	No	1	-	-	1	-	-	1	-	-
	Yes	**1.74**	**0.97-3.13**	**0.06**	1.35	0.65-2.80	0.42	**1.56**	**1.01-2.40**	**0.04**
Sleeps with sibling	No	1	-	-	1	-	-	1	-	-
	Yes	**2.73**	**1.39-5.31**	**0.00**	0.90	0.47-1.72	0.74	**1.52**	**0.97-2.36**	**0.06**
Microwave	No	1	-	-	1	-	-	1	-	-
	Yes	1.09	0.35-3.45	0.88	0.50	0.28-0.89	**0.01**	0.65	0.41-1.03	**0.06**
Paraffin	No	1	-	-	1	-	-	1	-	-
	Yes	-	-	-	1.27	0.60-2.71	0.53	1.17	0.58-2.37	0.67
Wood	No	1	-	-	1	-	-	1	-	-
	Yes	1	-	-	-	-	-	1.42	0.55-3.68	0.47
Gas	No	1	-	-	1	-	-	1	-	-
	Yes	0.41	0.09-2.01	0.27	-	-	-	0.26	0.04-1.53	0.14

[1]Other children staying with the child in the same household.

On the contrary, HIV exposure did not increase the risk of acquiring NVT at either TC Newman or Mbekweni (crude HR = 1.13, 95% CI = 0.76-1.71, p = 0.53). On the other hand, the presence of older siblings were significant drivers for acquisition of first episodes of NVT pneumococci at both TC Newman and Mbekweni, table 5.

Table S4.3: Cox proportional hazards model of time to first carriage acquisition of non-PCV-13 vaccine serotypes, stratified by site.

Risk factors		**TC Newman (n=59)**			**Mbekweni (n=77)**			**Crude (n=136)**		
		HR	**95% CI**	***p***	**HR**	**95% CI**	***p***	**HR**	**95% CI**	***p***
Gender	Female	1	-	-	1	-	-	1	-	-
	Male	1.05	0.80-1.39	0.70	0.96	0.76-1.22	0.76	1.00	0.84-1.20	0.970
Mode of delivery	Normal vaginal	1	-	-	1	-	-	1	-	-
	Normal vacuum	0.22	0.18-0.27	0.00	1	-	-	0.22	0.19-0.25	0.00
	Elective caesarean	0.95	0.73-1.24	0.71	1.15	0.86-1.53	0.34	0.58	0.26-1.28	0.18
	Emergency caesarean	0.94	0.67-1.31	0.71	1.04	0.79-1.38	0.78	1.00	0.83-1.22	0.99
HIV exposure	No	1	-	-	1	-	-	1	-	-
	Yes	1.25	0.85-1.83	0.25	0.88	0.69-1.13	0.32	0.95	0.78-1.17	0.66
Feed choice	Exclusive breastfed	1	-	-	1	-	-	1	-	-
	Mixed fed	1.13	0.87-1.46	0.37	0.90	0.61-1.32	0.58	1.48	1.12-1.96	0.00
	Never breastfed	1.02	0.75-1.39	0.88	0.78	0.57-1.06	0.12	0.91	0.73-1.12	0.37
Other Children[1]	No	1	-	-	1	-	-	1	-	-
	Yes	1.41	1.07-1.85	0.01	1.14	0.89-1.46	0.32	1.25	1.04-1.50	0.02
Premature	No	1	-	-	1	-	-	1	-	-
	Yes	1.13	0.83-1.54	0.43	0.81	0.51-1.26	0.35	0.95	0.72-1.24	0.69
Smoke exposure	No	1	-	-	1	-	-	1	-	-
	Yes	1.21	0.82-1.79	0.35	1.09	0.86-1.39	0.46	1.09	0.90-1.31	0.40
Sleeps with Mother	No	1	-	-	1	-	-	1	-	-
	Yes	1.10	0.55-2.19	0.78	0.70	0.52-0.94	0.01	0.80	0.60-1.06	0.12
Sleeps with Father	No	1	-	-	1	-	-	1	-	-
	Yes	1.09	0.83-1.44	0.55	1.09	0.86-1.39	0.48	1.08	0.90-1.28	0.41

Sleeps with sibling	No	1	-	-	1	-	-	1	-	-
	Yes	1.37	1.03-1.82	0.03	1.13	0.88-1.44	0.33	1.22	1.02-1.46	0.03
Microwave	No	1	-	-	1	-	-	1	-	-
	Yes	1.01	0.59-1.72	0.98	0.80	0.62-1.04	0.09	0.83	0.67-1.04	0.11
Paraffin	No	1	-	-	1	-	-	1	-	-
	Yes				1.07	0.83-1.39	0.61	1.09	0.85-1.39	0.50
Wood	No	1	-	-	1	-	-	1	-	-
	Yes	1.36	1.00-1.85	0.05	1	-	-	1.32	0.99-1.74	0.05
Gas	No	1	-	-	1	-	-	1	-	-
	Yes	0.57	0.29-1.11	0.09	0.56	0.15-2.06	0.38	0.55	0.29-1.06	0.07

[1]Other children staying with the child in the same household.

4.3 Methodology and mean registered curves for all other serotypes

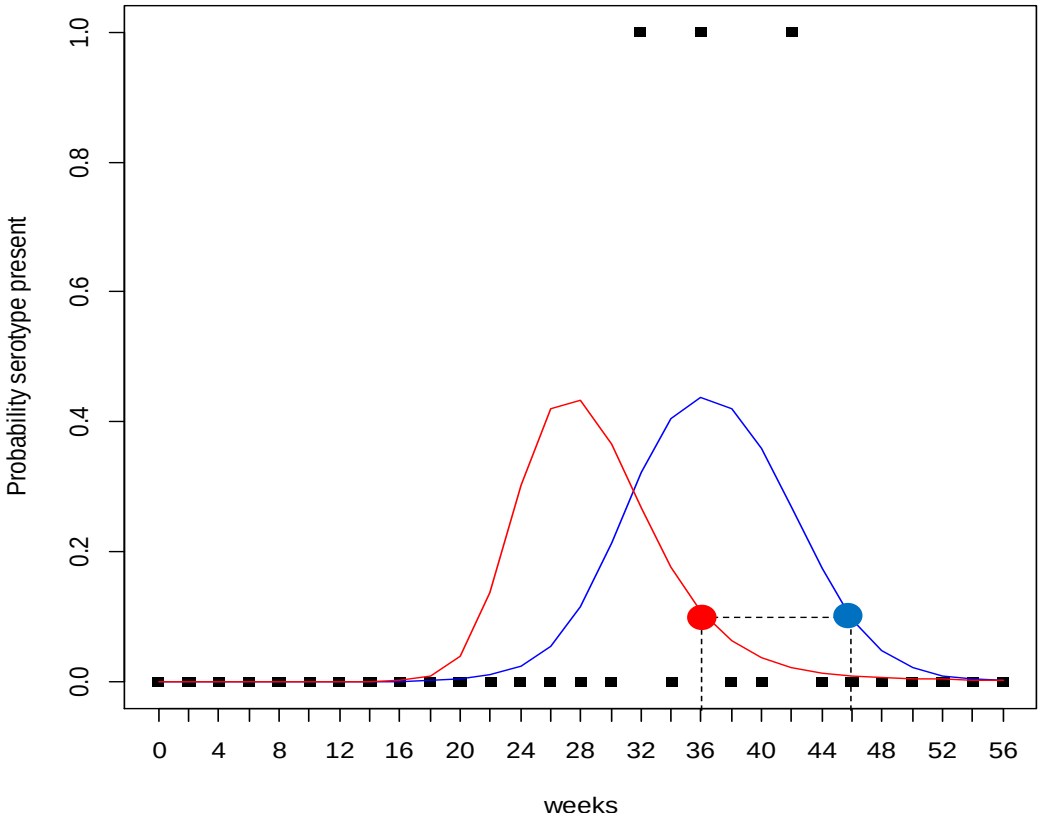

Figure S4.1: Landmark curve registration for non-parametric logistic regression fit.

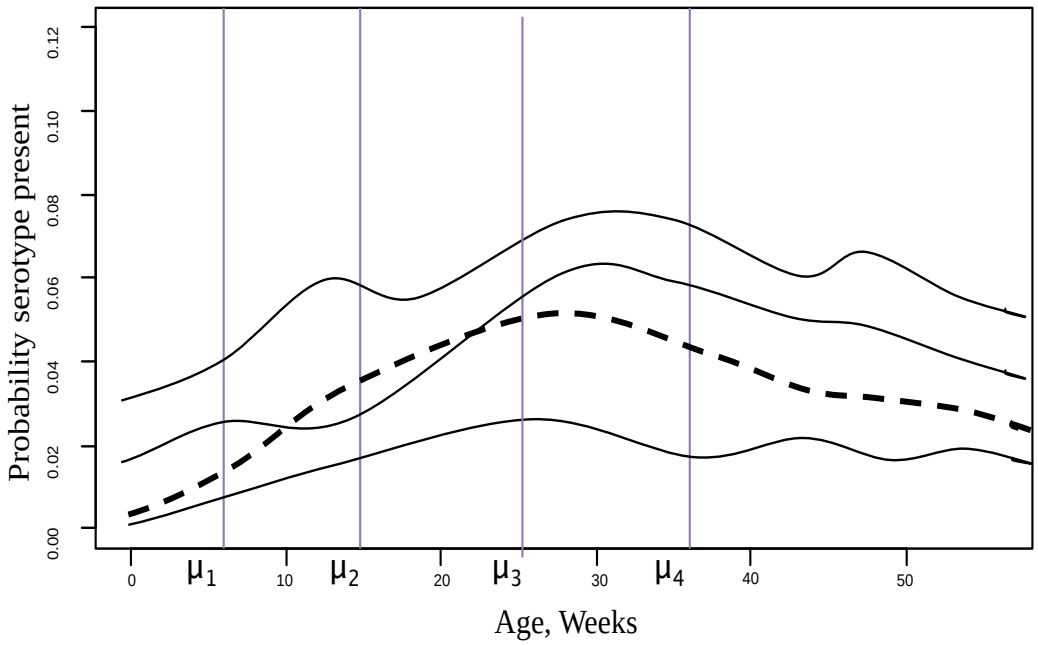

Figure S4.2: Mean registered curves for each infant for a single serotype. As an illustration, only three children carrying one specific serotype are represented, black solid lines and the mean registered curve is shown by the black dotted lines. Therefore the mean estimated probability that the serotype is present at each time point is shown by μ_{n+1}.

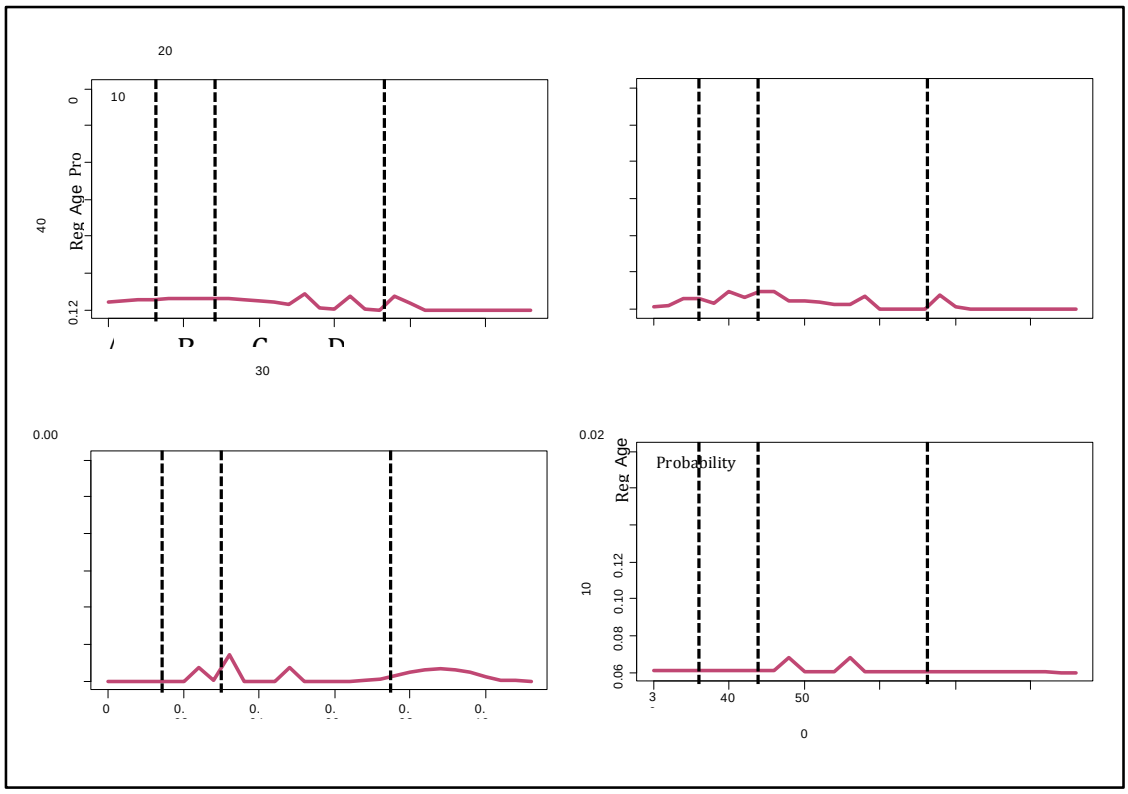

Figure S4.3: Mean registered curves showing the mucosal carriage of individual PCV-13 serotypes with 95% upper and lower confidence intervals (dotted red lines). A = Serotype 6B, B = Serotype 14, C = Serotype 3 and D = Serotype 4

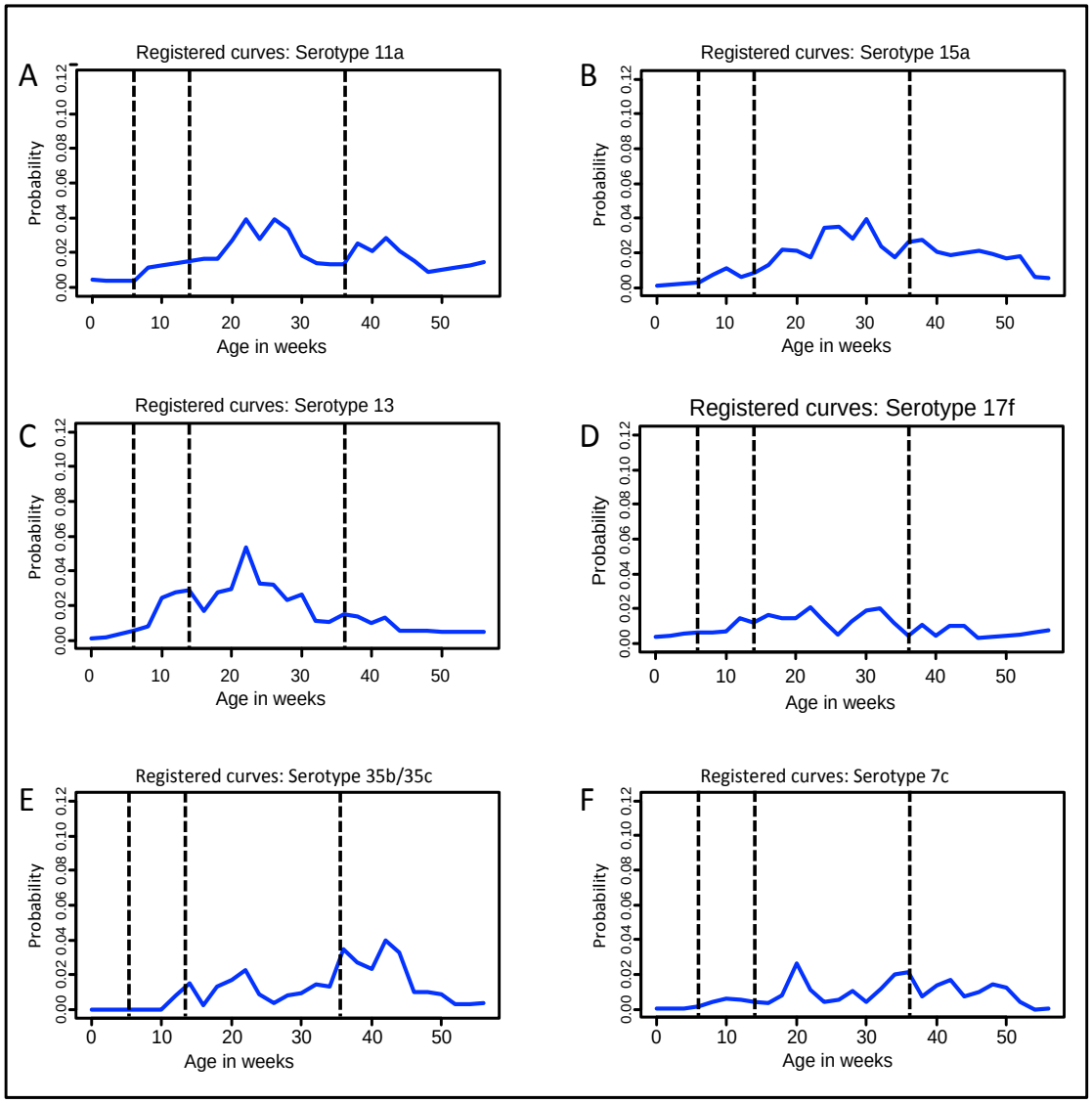

Figure S4.3: Mean registered curves for the additional non-PCV-13 vaccine serotypes. A= Serotype 11A, B= Serotype 15A, C= Serotype 13, D= Serotype 17F, E= Serotype 35B/35C and F= Serotype 7C. Dotted verticle lines represents the 1st, 2nd and 3rd vaccination dates.

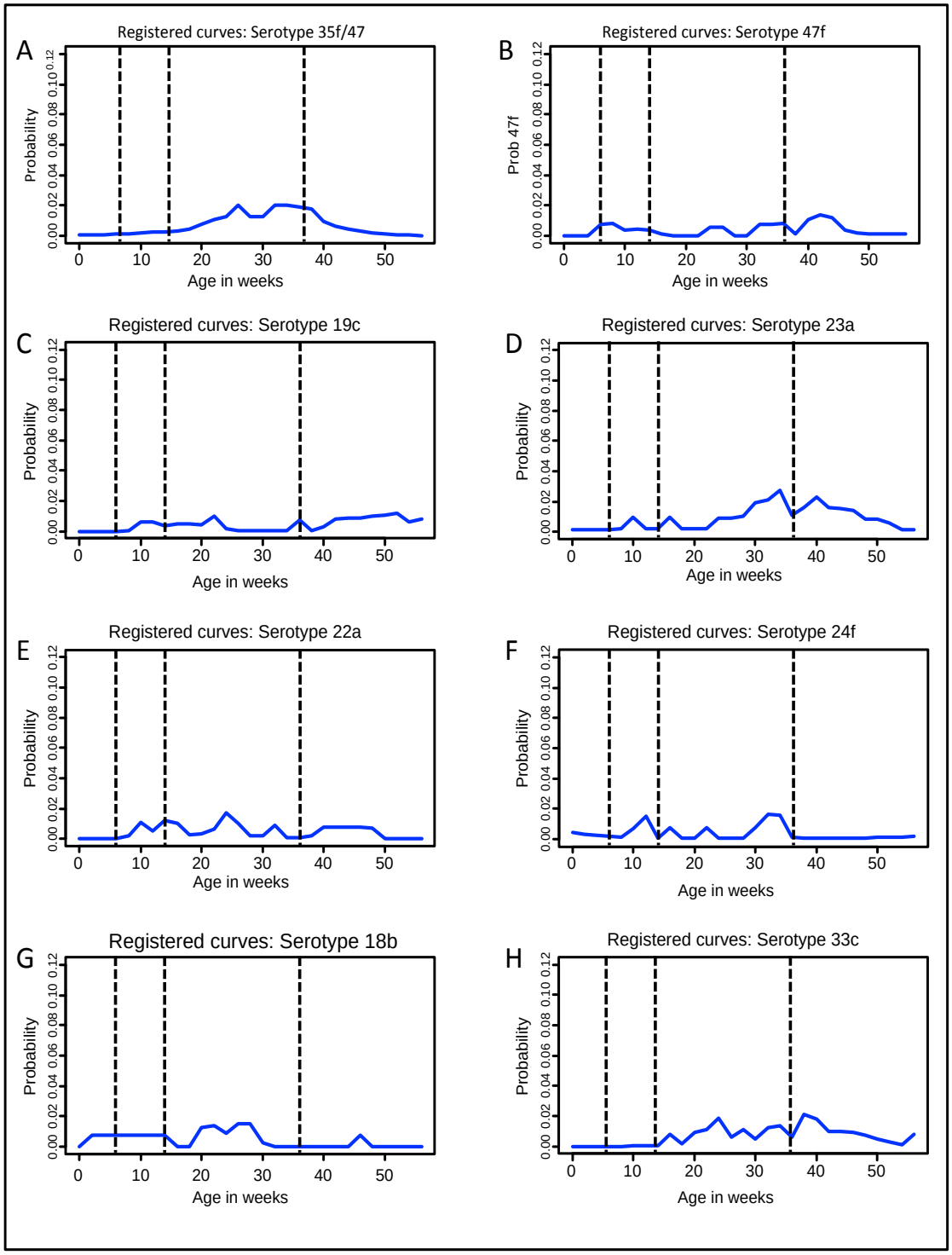

Figure S4.4: Mean registered curves for the less commonly encountered non-PCV-13 vaccine serotypes. A= Serotype 35F/47, B= Serotype 47F, C= Serotype 19C, D= Serotype 23A, E= Serotype 22A, F= Serotype 24F, G= Serotype 18B and E= Serotype 33C. Dotted verticle lines represents the 1st, 2nd and 3rd vaccination dates.

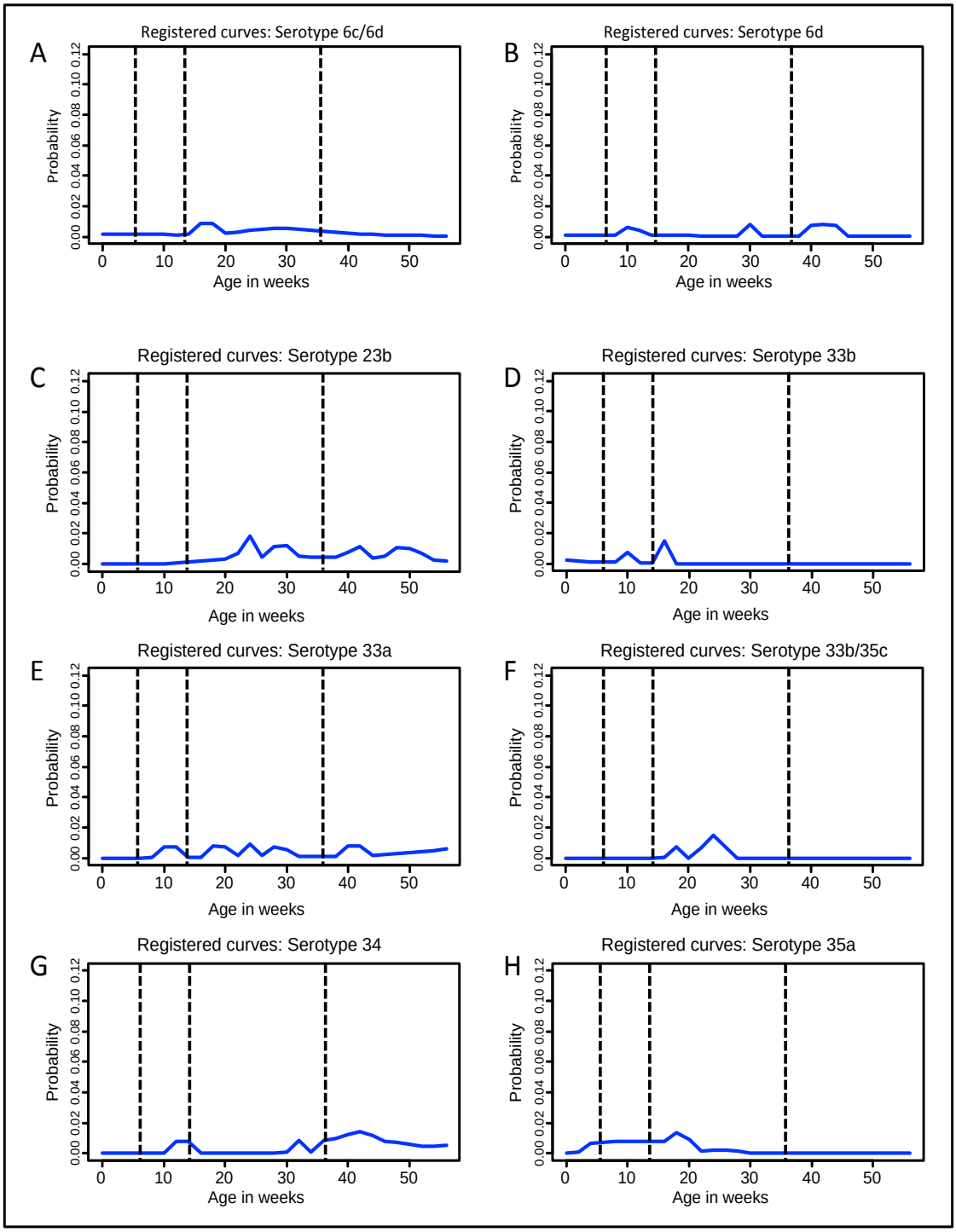

Figure S4.5: Mean registered curves for the less commonly encountered non-PCV-13 vaccine serotypes.

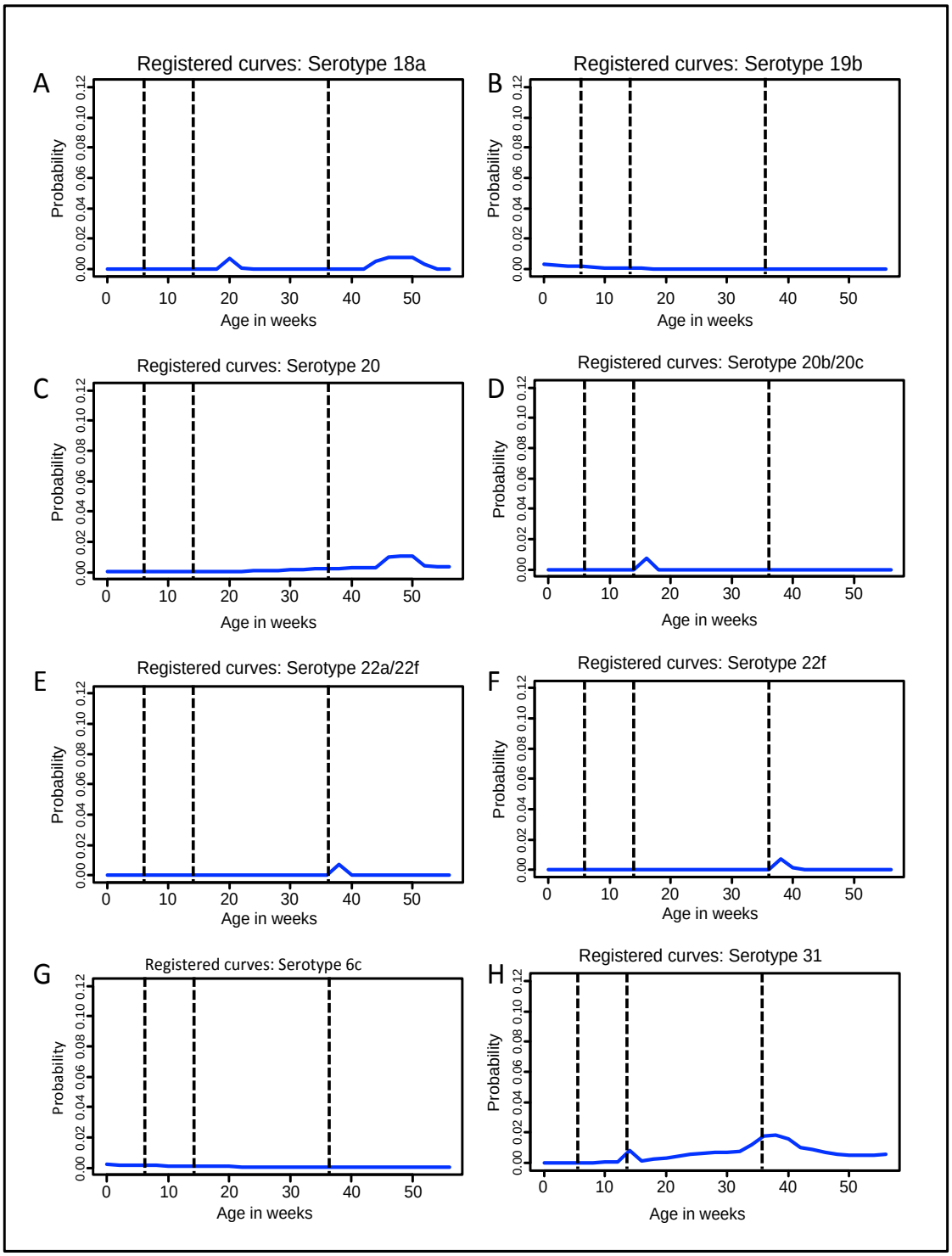

Figure S4.6: Mean registered curves for the less commonly encountered non-PCV-13 vaccine serotypes.

APPENDIX 5: Chapter 6 supplementary tables and figures

Table S5.1: Summary paired of pathogen co-oncurrence in the **TB** category

	FLUA	RHINO	FLUB	COR43	RSVAB	CMV	AV	EVPV	PARA1	HMPVAB	MPNEU	HBOV	S_AUR	C_PNEU	S_PNEU	HIB	PCP	MORAX	FLUC	HAEINF
Rhino	1																			
FluB	1	1																		
Cor43	1	0.66	1																	
RSVAB	1	1	1	1																
CMV	0.10	1	1	1	1															
AV	1	1	1	1	1	0.26														
EVPV	1	1	1	1	1	1	1													
Para1	1	1	1	1	1	1	1	1												
HMPVAB	0.21	0.79	1	0.72	1	0.11	0.53	1	1											
Mpneu	1	1	1	1	0.06	1	1	1	1	1										
Hbov	1	0.18	1	1	1	1	1	1	1	1	1									
S_aur	1	0.37	1	1	0.21	1	1	0.21	1	0.85	0.38	0.21								
C_pneu	1	0.58	1	1	1	0.33	1	1	1	1	1	1	0.64							
S_pneu	0.54	0.42	0.54	0.23	1	0.58	0.91	1	1	0.29	0.81	0.54	0.61	1						
HiB	1	0.21	1	0.58	1	1	0.41	1	1	1	1	1	1	0.51	0.60					
PCP	1	0.46	1	0.40	0.09	1	1	1	1	1	0.17	1	0.11	1	0.57	1				
Morax	0.72	0.47	0.72	0.18	0.73	0.38	0.82	1	1	0.37	0.53	1	0.93	0.95	**0.00**	0.58	0.37			
FluC	1	0.74	1	0.66	1	1	0.46	1	1	0.79	0.34	1	1	0.58	0.95	0.22	1	0.81		
Haeinf	0.24	0.13	1	1	1	**0.01**	0.58	1	1	0.21	1	0.24	0.56	0.24	0.17	0.77	1	0.27	1	
Bord	1	0.46	1	0.40	1	1	1	1	1	1	1	1	1	0.33	0.91	0.06	1	0.37	0.08	0.57

**The diagonal line represents comparison to self which empirically has an achieved level of significance (ASL) of p=1. Significant concurrent relationships are highlighted in bold cyan.*

Table S5.2: Summary of achieved significance level for all paired pathogen co-occurrences in the **non-TB** category

	FLUA	RHIN	FLUB	P3	P2	P4	C63	C43	HKU	RSV	CMV	AV	EPV	P1	HMP	MP	HBV	SA	CPN	SPN	HIB	PCP	MX	FLC	HNF
Rhino	1																								
FluB	1	1																							
Para3	1	0.14	1																						
Para2	1	0.13	1	1																					
Para4	1	1	1	1	1																				
Cor63	1	1	1	1	1	1																			
Cor43	1	1	1	1	1	1	1																		
CorHKU	1	1	1	1	1	1	1	1																	
RSVAB	1	1	1	1	1	1	1	1	1																
CMV	1	1	1	1	1	1	1	0.17	0.10	1															
AV	1	0.61	1	1	1	1	1	1	1	1	0.46														
EVPV	1	0.18	0.43	1	1	1	1	1	1	1	0.68	0.17													
Para1	1	1	1	1	1	1	0.02	1	1	1	1	1	1												
HMPVAB	1	0.26	0.49	1	1	1	1	0.28	1	1	0.75	1	0.21	1											
Mpneu	0.17	0.31	1	0.10	1	1	1	0.18	1	1	1	0.45	1	1	0.35										
Hbov	0.04	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1									
S_aur	1	0.98	1	1	1	0.26	1	0.45	1	1	0.11	1	0.11	1	0.43	1	0.44								
C_pneu	1	1	1	1	1	1	1	1	1	1	1	1	0.13	1	1	1	1	1							
S_pneu	0.64	0.56	0.87	1	1	1	1	1	0.40	1	0.39	0.77	0.71	0.64	0.20	0.69	0.64	0.34	1						
HiB	1	1	1	1	1	1	1	1	1	1	0.68	0.03	0.43	1	1	0.27	1	0.03	1	0.46					
PCP	1	0.80	1	1	1	1	1	1	0.12	1	0.24	1	0.76	1	0.82	0.65	1	0.02	1	0.14	0.00				
Morax	0.89	0.36	0.59	0.66	1	1	0.67	0.44	0.66	0.74	0.45	0.99	0.46	0.43	0.54	0.17	1	0.86	0.66	0.18	0.89	0.79			
FluC	1	1	1	1	1	1	1	1	1	0.14	1	1	1	1	0.49	0.34	1	1	1	1	0.42	0.40	0.58		
Haeinf	0.11	0.84	0.41	1	1	1	1	0.11	0.34	0.71	0.08	0.32	0.79	0.57	0.46	0.27	0.56	0.94	0.34	0.17	0.55	0.21	0.28	1	
Bord	1	1	0.14	1	1	1	1	1	1	1	1	1	0.34	1	0.39	1	1	1	1	0.78	1	1	0.29	0.14	0.26

**The diagonal line represents comparison to self which empirically has an achieved level of significance (ASL) of p=1. Significant concurrent relationships are highlighted in bold cyan.*

Table S5.3: Summary of achieved significance level for all paired pathogen co-occurrences*.

	FluA	Rh	FluB	P3	P2	P4	C29	C63	C43	HKU	RSV	CV	AV	EPV	P1	HmPV	Mp	Hbov	Sa	Cp	Sp	HiB	PCP	Mx	FluC	Hif
Rhinovirus (Rh)	1																									
Influenza B (FluB)	1	1																								
Parainfuenza 3 (P3)	1	1	1																							
Parainfuenza 2 (P2)	1	1	1	1																						
Parainfuenza 4 (P4)	1	1	1	1	1																					
Coronavirus 229 (C229)	1	1	1	1	1	1																				
Coronavirus 63 (C63)	1	1	1	1	1	1	1																			
Coronavirus 43 (C43)	1	1	1	1	1	1	1	1																		
Coronavirus HKU	1	1	1	1	1	1	1	1	1																	
RSV A/B[a] (RSV)	1	1	1	1	1	1	1	1	1	1																
Cytomegalovirus (CV)	1	1	1	1	1	1	1	1	1	1	1															
Adenovirus (AV)	1	1	1	1	1	1	1	1	1	1	1	1														
Enterovirus (EPV)	1	1	1	1	1	1	1	1	1	1	1	1	1													
Parainfluenza 1 (P1)	1	1	1	1	1	1	1	1	1	1	1	1	1	1												
HmPV A/B[b] (HmPV)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											
M. pneumoniae (Mp)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1										
Bocavirus (Hbv)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1									
S. aureus (Sa)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1								
C.pneumoniae (Cp)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1							
S.pneumoniae (Sp)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
H.influenzae B (HiB)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1					
P. jirovecii	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.03	1	0.76	1				
M. cartaarhalis (Mx)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Influenza C(FluC)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.49	1	1	1		
H. influenzae spp (Hif)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
B. pertussis (Bd)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.07	1	1	1	1	1

**The pathogen abbreviations names in the top row are derived from the first column. The diagonal line represents comparison to self which empirically has an ac hieved level of significance (ASL) of p=1, the table reports ASL for all concurrent pathogen pairs irrespective of the TB category. Significant concurrent relationships are highlighted in bold red. All comparisons have been adjusted for false discovery multiple comparisons.* [a]*RSV A/B = Respiratory Syncyntial virus A and B*, [b]HmPV A/B = Human Metapneumovirus A and B

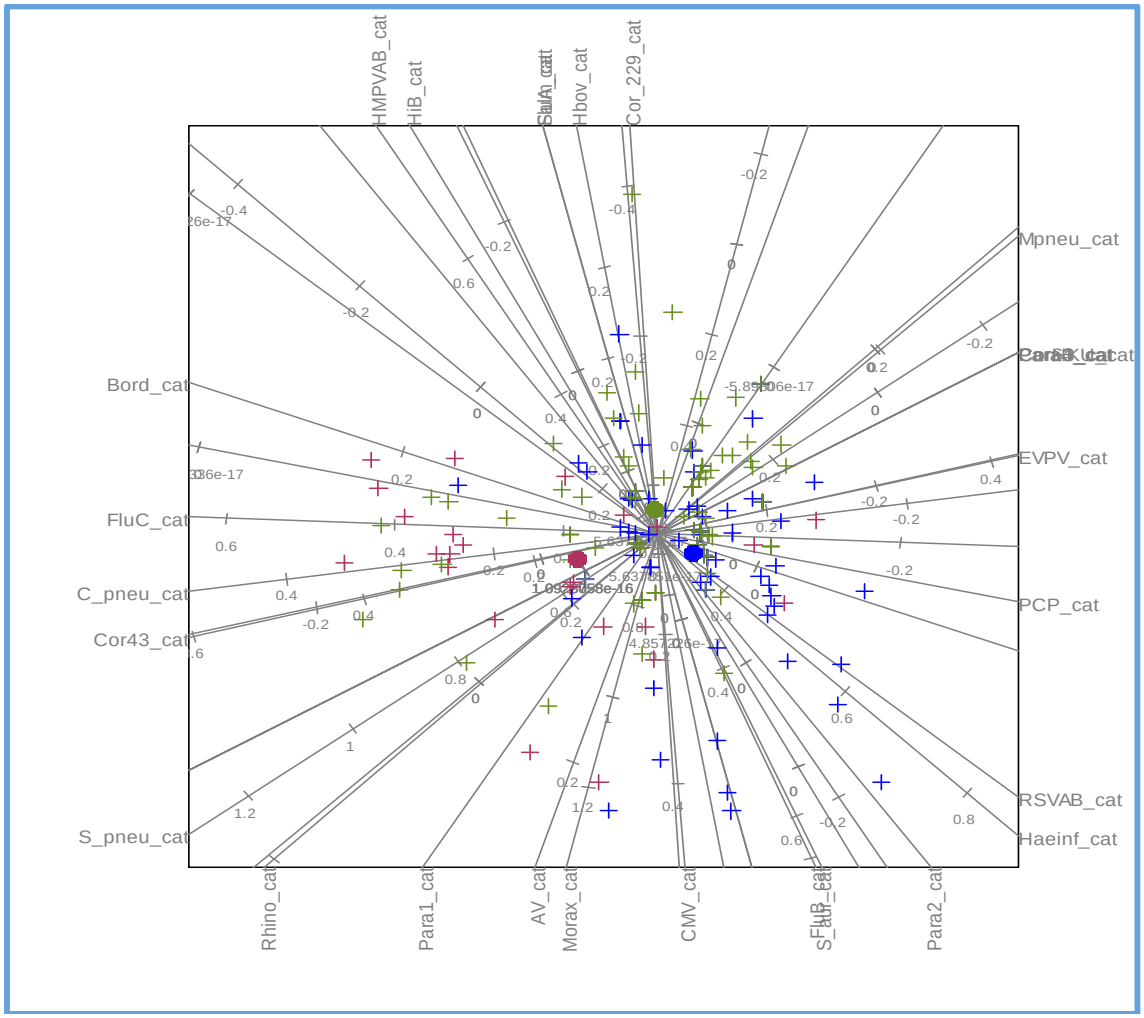

Figure S5.1: Cononical variate analysis (CVA) biplot showing the presence or absense of respiratory pathogens in the definite TB (Red squares), non TB (Blue squares) and possible TB (Green squares) groups.

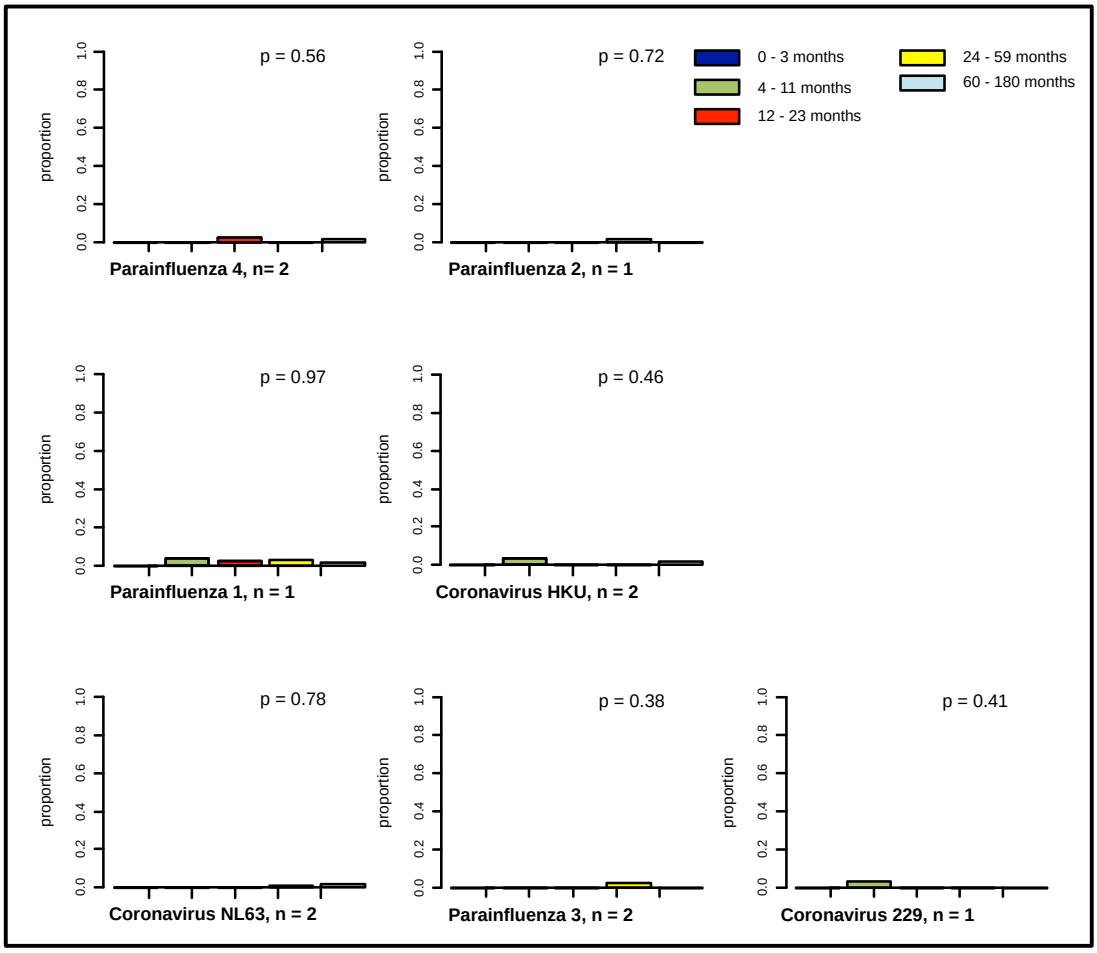

Figure S5.2: Age-dependant viral pathogen distribution among different children's age groups. Note: Age groups 0–3 months, 4–11 months, 12–23 months, 24–59 months, and 60–180 months.

	POS	NEG	111142	111162	111299	111315	111374	111452	111468	111557	112678	308861	310159	512197
Internal Control		32.12	34.95	33.70	33.99	34.10	34.52	33.51	35.10	32.88	33.29	34.52	32.42	33.03
Influenza A	31.40													
Rhinovirus	28.11							39.24	38.42					
Influenza B	29.05													
Parainfluenza 3	29.28													
Parainfluenza 2	32.26													
Parainfluenza 4	29.56													
Coronavirus 229	32.19													
Coronavirus 63	29.11													
Coronavirus hku	27.81													
Coronavirus 43	30.63		21.46											
Parainfluenza 1	29.40													
Metapneumovirus A and B	27.72		32.27					28.31	32.59	34.35	36.70			
Bocavirus	29.42													
M_ pneumoniae	26.88		23.78						35.31					
Respiratory syncyncial virus A and B	29.27													
Cytomegalovirus	27.40		32.11					28.09	32.01	33.61				
Enterovirus/Parechovirus	29.22													
Adenovirus	26.67		23.28						36.64					
S_ aureus	28.17				27.88				32.16		23.70		27.02	
C_ pneumoniae	28.42													
H. influenzae B	29.16													
S_ pneumoniae	28.90		26.35			21.00	23.25	29.25				27.41		
Pneumocystis jirovecii	28.01			30.54							31.29			
Legionella spp	27.27		36.92			35.15	35.21	35.94	38.05	36.73		36.88		
k_ pneumoniae	27.16				34.38	34.54	34.06			35.29	34.47	36.48	34.80	36.34
Salmonella spp	28.97													
M_catarrhalis	29.41		25.30	25.27		19.11	25.04	23.46	25.87	25.35		28.09		
Influenza C	28.51													
Bordetella pertussis	26.75													
H. influenzae spp	28.21		23.98			20.16	20.78	24.67				25.08		

Figure S5.3: Screen shot of the FTDResp33 Analyser results output. The first column contains the names of each pathogens included in the FTD Ressp33 assay, and the adjucent rows iincludes the cycle threhold value for each pathogen detected for all the 12 clinical samples, column 4 – 15. Empty cells represents negative samples.

5.4 How to run the FTD Resp33 qPCR Analyser program

In order to FTD results, downloaded the FTD Resp33 qPCR Analyser program at http://www.gematics.com/analyser.html, which is stored as a zip folder. After extracting the program to a specified location:

1) Users of either MAC or windows operating system must launch the program by double clicking on the analyser.jar file located in the deployAnalyser directory.
2) To import the data files click on "File" and then select one of two options. For relative quantification select the first option and for absolute quantification select the second option. The relative quantification approach rearranges the Cq values obtained from the CFX realtime machine into a table that
3) If you are performing relative quantification, you will need to specify a set of gradients and intercepts for the calculation. Click on the "Open constants file"

button. Each sheet in the file contains a set of slopes and gradients. You will be able to select which set to use when the calculation is performed.

4) When performing absolute quantification, the absolute numbers will be read in from the input files.
5) When the calculation is complete, the "View excel document" button will appear. Click on this to view the results.
6) To change the order in which the rows are displayed, click on the "Open names file" button. Here you can specify the full name for each pathogen and the order they are displayed in.run.command for users.
7) Click on “File” and then “Open directory”
8) Now navigate to a secified directory that contains the AB1 files which have been labelled using a set convention (See under rebuilding the database for the naming format).
9) Once you have selected a directory, click on the “Open” button
10) A window will open to notify the user if there are any errors in file processing which can include sequences that for below the 0.05 PHRED cutoff or sequences that are duplicated or have a missing compliment sequence.
11) If this QC step is resolved, click on “Proceed anyway”
12) The program will then find all the alignments for the input files.
13) Once the alignment is complete, a report window will open showing the results.