Understanding the Dynamics and Spread of Pneumococcal Infection From Healthy Carriage to Pneumonia and Invasive Disease, in Kilifi, Kenya

Thesis

UNRESTRICTED

Understanding the dynamics and spread of pneumococcal infection from healthy carriage to pneumonia and invasive disease, in Kilifi, Kenya

---

PhD Thesis for the Open University

Life and Biomolecular Sciences

ARC: KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

Susan Claire Morpeth, MBChB, FRACP, FRCPA, DTM&H

November 2014

Director of Studies: Prof Anthony Scott, University of Oxford and

London School of Hygiene & Tropical Medicine

Co-supervisors: Prof David Murdoch, University of Otago

Prof James Nokes, Warwick University

Date of Submission: 1 November 2014

Date of Award: 10 April 2015

Abstract

Streptococcus pneumoniae, the pneumococcus, is an important pathogen globally, especially among children <5 years old. Pneumococcus colonises the human nasopharynx, and carriage is a prerequisite to disease. Understanding the concentration of pneumococcal colonisation may help us to better understand the pathogenesis of pneumococcal disease.

We hypothesized that If increased nasopharyngeal pneumococcal concentration is important in the pathogenesis of pneumonia then the concentration would be lowest among well children, intermediate among children with symptomatic upper respiratory tract infection (URTI) and highest among children admitted to hospital with WHO-defined pneumonia. We found that in fact, the nasopharyngeal pneumococcal concentration was of similar distribution among all three groups of children, and was highest among the group with URTI.

We studied the nasopharyngeal concentration of pneumococcus among children <5 years old before, during and after episodes of symptomatic URTI with respiratory syncytial virus (RSV) or rhinovirus. Nasopharyngeal pneumococcal concentration increased 4-fold with onset of the viral infection, which may contribute to risk.

We examined the response of nasopharyngeal pneumococcal concentration to vaccination with the 10-valent pneumococcal conjugate vaccine (PCV-10) among children aged 12-23 months. Vaccine-type (VT) carriage concentration did not change, while non-vaccine-type (NVT) concentration increased over six months post vaccination, which may have been a vaccine effect on VT relative to NVT concentration. There were large

differences in serotype-specific pneumococcal concentration within and between individuals.

We investigated the performance of quantitative *lytA* PCR (qPCR) on blood samples in the diagnosis of invasive pneumococcal disease (IPD) among children ≤13 years old at Kilifi County Hospital, and among children 1-59 months old in a seven-country case-control study of pneumonia aetiology. *lytA* qPCR was not specific for the diagnosis of IPD, being positive among 5% of control participants. Pneumococcal DNA is more commonly found in the bloodstream of children in low and lower-middle income countries than has been appreciated.

Studying nasopharyngeal concentration of pneumococcus demonstrated why it is difficult to diagnose pneumococcal disease among young children in high-burden countries. Total nasopharyngeal concentration of pneumococcus is similar in well children as in disease. Serotype-specific concentration varies widely within and between children. Pneumococcal bloodstream invasion appears to be a frequent occurrence, without always leading to IPD.

Acknowledgements

I would like to thank my supervisors, Prof Anthony Scott, Prof David Murdoch and Prof James Nokes for their scientific mentorship, support and encouragement. I also thank Prof Tom Williams, my third-party monitor, for his advice and support.

Collaborators Patrick Munywoki, Laura Hammitt, Jason Hinds, and the PERCH Study Group; thank you very much for allowing me to work with samples from your studies and for helpful discussions of analysis plans and of my findings. I am particularly grateful for Patrick's help with STATA coding.

Nicole Wolter and Anne von Gottberg at the NICD in Johannesburg, thank you for hosting me and teaching me.

Jim Huggett at the LGC, thank you for suggesting the human DNA quantification method and for helpful discussions about making standards for quantitative PCR.

John Ojal, Greg Fegan, Michael Ooko, Alice Kamau and Chizi Mwango, thank you all for statistical and data management support.

Thank you to all the laboratory staff and scientists at the KEMRI-Wellcome Trust Research Programme, particularly; Sammy Nyongesa, Jane Jomo, Phelgona Awuor, Anne Karanu, Shadrack Mutua, Anne Amulele, James Otieno, Angela Karani and the pneumococcal research team, Salim Mwarumba and the bacteriology team, Clayton Onyango, Anne Bett and the virology team, Alex Macharia and the TW team, Brett Lowe, Moses Mosobo and the laboratory support staff.

Alex Maina, thank you for library support. Tahreni Bwaanali, Dorothy Tole, Elinah Kaulu and Liz Murabu, thank you for administrative support.

Juliet Otieno and the Kilifi PERCH team, and the study teams for WAIFW and PRISM, thank you very much.

Margaret Mackinnon, Ifedayo Adetifa, Sarah Atkinson, Sonal Henson, Betty Owor, Tony Etyang, Faith Osier, thank you all for your time and patience in helping improve this work.

To all the patients, caregivers and staff at Kilifi County Hospital and study participants in the community, asanteni sana.

For my beloved father; Gavin Hector Morpeth, 1944 – 2013

and

for my daughter, Malaika Sage Morpeth, born in November 2013, the light of my life

Table of Contents

Table of figures

Table of tables

List of Abbreviations

ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
CbpA	Choline binding protein A
CbpE	Choline binding protein E
CDC	Centers for Disease Control and Prevention
CFU	Colony forming units
CO2	Carbon dioxide
Ct	Cycle threshold
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FTD	Fast-Track Diagnostics
HIV	Human immunodeficiency virus
hMPV	Human metapneumovirus
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL	Interleukin
IPD	Invasive pneumococcal disease
IQR	Interquartile range
IRAK-4	interleukin-1 receptor associated kinase-4
KCH	Kilifi County Hospital
KEMRI	Kenya Medical Research Institute
KHDSS	Kilifi Health and Demographic Survey System
KWTRP	KEMRI-Wellcome Trust Research Programme
MIF	Macrophage migration inhibitory factor
MLST	Multi-locus sequence typing
MUAC	Mid-upper arm circumference
MyD88	Myeloid differentiation factor
NAATs	Nucleic acid amplification tests
NCBI	National Center for Biotechnology Information
NF-KB	Nuclear factor-kappaB
NICD	National Institute for Communicable Diseases
NO	Nitric oxide
NP/OP	Nasopharyngeal/oropharyngeal

NPFS	Nasopharyngeal flocked swab
NT	Not typable
NVT	Non-vaccine type
PAF	Platelet-activating factor
PAFR	Platelet-activating factor receptor
PavA	Pneumococcal adherence and virulence factor A
PavB	Pneumococcal adherence and virulence factor B
PBP	Penicillin binding protein
PBS	Phosphate-buffered saline
PCV	Pneumococcal conjugate vaccine
PCV-10	10-valent pneumococcal conjugate vaccine
PCVIS	Pneumococcal Conjugate Vaccine Impact Studies
PERCH	Pneumonia Etiology Research for Child Health
PfbA	Plasminogen and fibronectin binding protein A
PfbE	Plasminogen and fibronectin binding protein E
pIgR	Polymeric immunoglobulin receptor
PRISM	PCV-10 Reactogenicity, Immunogenicity and Safety study in Malindi
PsaA	Pneumococcal surface adhesin A
PspC	Pneumococcal surface protein C
qPCR	Quantitative PCR
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
ROC	Receiver-operating characteristic
RSV	Respiratory syncitial virus
SFTPC	Surfactant protein C
STGG	Skim-milk tryptone glucose glycerol
TGF	Transforming growth factor
Th17	IL-17-secreting CD4+ T-cell
TNA	Total nucleic acid
TNF-α	Tumour necrosis factor alpha
URTI	Upper respiratory tract infection
UTM	Universal transport medium
VT	Vaccine type
VTM	Viral transport medium
WAIFW	Who Acquires Infection from Whom
WHO	World Health Organisation

Chapter One

1 Thesis Outline

1.1 Introduction

Pneumonia is the leading cause of death among children <5 years of age worldwide, and in Africa.[1] *Streptococcus pneumoniae*, as well as being a leading cause of pneumonia,[2] causes invasive pneumococcal diseases (IPD) such as sepsis and meningitis, with a high case fatality. The reservoir of *S. pneumoniae*, the pneumococcus, is carriage in the human nasopharynx, particularly the nasopharynges of young children. Diagnosis of pneumococcal pneumonia and IPD is difficult, especially among young children in low and lower-middle income countries where carriage prevalence is highest and the burden of disease is greatest.

1.2 Research Question

To better understand the relationship between healthy carriage of pneumococcus and pneumococcal pneumonia/invasive pneumococcal disease, and how this relationship affects our ability to diagnose pneumococcal disease in young children, using quantitative PCR to measure pneumococcal concentration.

1.3 Specific Objectives

1. To quantify the nasopharyngeal concentration of pneumococcus in different disease states to understand the constraints on the diagnosis of pneumococcal pneumonia in young children.

    a. Is nasopharyngeal pneumococcal concentration greater in children with pneumonia than in children with symptoms of upper respiratory tract infection (URTI) symptoms or in healthy children?

2. To describe the dynamics of pneumococcal concentration in the nasopharynx during symptomatic URTI with respiratory syncytial virus (RSV), or rhinovirus in children <5 years of age.

    a. Do viral infections lead to an increase in nasopharyngeal pneumococcal concentration?

    b. If viral infections increase nasopharyngeal pneumococcal concentration, does this return to baseline after the viral URTI?

3. To describe the dynamics of pneumococcal concentration in the nasopharynx after immunisation with 10-valent pneumococcal conjugate vaccine (PCV-10).

    a. Does nasopharyngeal total pneumococcal concentration change after PCV-10?

    b. Does vaccine-type pneumococcal concentration decrease after PCV-10?

    c. Does non-vaccine-type pneumococcal concentration increase after PCV-10?

4. To investigate the performance of *lytA* qPCR in the diagnosis of invasive pneumococcal disease using blood samples, in children in low and lower-middle income countries.

1.4 Overview and context of the thesis

The pneumococcal pathogenic process begins with transmission and acquisition of nasopharyngeal carriage of pneumococcus, followed by amplification, mucosal disease and finally invasive disease. Nasopharyngeal carriage of pneumococcus is thus a prerequisite for disease, but does not always lead to disease. Pneumococcal carriage is very common, with highest prevalence among young children. Diagnosis of pneumococcal disease is particularly difficult in young children in developing countries, precisely the group with the highest burden of disease where diagnosis is important.

The thesis begins with the study of nasopharyngeal pneumococcal concentrations in health and disease, followed by an examination of the impact of symptomatic respiratory virus infection on nasopharyngeal pneumococcal concentration, in young children. Next, the effect of pneumococcal conjugate vaccination on nasopharyngeal pneumococcal concentration is studied. Finally the diagnosis of invasive pneumococcal disease is studied.

Nasopharyngeal pneumococcal density and pneumococcal concentration are both terms used in the published literature. Terms such as nasopharyngeal pneumococcal load or pneumococcal burden are also used. Density is a measure of mass per unit volume; in this case the mass of pneumococcus per unit volume in the nasopharyngeal secretions adherent to the posterior nasopharyngeal mucosae. In estimating the pneumococcal

density in the nasopharynx all that we can actually measure is the concentration of pneumococcus in a swab sample. For work performed in this thesis, the term nasopharyngeal pneumococcal concentration will be used. Methods exploring the estimation of nasopharyngeal pneumococcal concentration from a nasopharyngeal swab are detailed in the methodology section.

The research objectives in this thesis were nested within studies of respiratory and pneumococcal disease that took place at the Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme (KWTRP) in Kilifi, Kenya, and within a larger multi-site pneumonia study. At Kilifi County Hospital (KCH), all paediatric medical admissions (ie. all admissions excluding accidents, surgical cases, poisoning and fresh burns) were invited to consent for collection of blood for invasive bacterial disease surveillance as part of the Pneumococcal Conjugate Vaccine Impact Studies (PCVIS),[3] and Kilifi genome-wide association studies. The Pneumonia Etiology Research for Child Health (PERCH) project was a multi-site case-control study of the causes of pneumonia among children aged 1-59 months that took place in Kilifi, Kenya, as well as in the Gambia, Mali, Zambia, South Africa, Thailand and Bangladesh over 2011-2014.[4] Respiratory virus surveillance and studies of the aetiology of pneumonia have been carried out longitudinally at KCH [5-7] and data and sample collection for the PERCH study were introduced in a stepwise fashion in this context.

1.4.1 Objective One

The first objective asked whether nasopharyngeal pneumococcal concentration was greater in children with pneumonia than children with URTI symptoms or healthy children, in order to better understand the constraints on pneumococcal pneumonia diagnosis in young children. This objective used samples from the PERCH project.

1.4.2 Objective Two

The second objective investigated whether pneumococcal concentration increased in response to a RSV or rhinovirus URTI in young children. This was a sub-study of a household RSV transmission study that took place in Kilifi County in 2009-2010.[8,9] In the parent study, nasopharyngeal swabs were collected from all members of 47 households with at least one infant and a sibling, twice weekly for six months. A multiplex PCR was used to detect the presence of respiratory viruses.

1.4.3 Objective Three

The third objective examined the effect of PCV-10 on serotype-specific pneumococcal concentration in the nasopharynx. This was a sub-study of a randomized controlled trial of the immunogenicity, reactogenicity and impact on carriage of PCV-10 among children aged 12-59 months that took place in Malindi, Kenya (PRISM).[10]

1.4.4 Objective Four

The fourth objective studied the performance of *lytA* qPCR from paediatric blood specimens in the diagnosis of IPD. This study was carried out on two sample sets, the first from ongoing paediatric invasive bacterial disease surveillance at KCH carried out at the KWTRP, which included participants in the Kilifi arm of the PERCH project. The second data set consisted of cases and controls from all sites participating in the PERCH project.

1.4.5 Declaration of the author's role

The author of this thesis conceived the idea for this work as a whole and for each of the objectives. She was the laboratory lead for the PERCH project in Kilifi and worked closely with the PERCH core laboratory lead, Professor David Murdoch, in setting up and

validating laboratory assays for the PERCH project. Likewise, she led the KWTRP microbiology laboratory for the PRISM study and for the PCVIS project.

Method validation: The author designed and performed most of the work herself, or in some instances supervised a laboratory team in the performance of experiments designed by the author. In the comparison of DNA extraction instruments, experiments using the ABI PRISM 6100 were performed by the Kilifi genome-wide association studies laboratory team and were co-supervised by the author. Experiments designed by the author on DNA extraction methods for nasopharyngeal swabs in skim-milk tryptone glucose glycerol (STGG), and the effect of transport time of swabs in STGG on the measured pneumococcal concentration, were performed by an intern supervised by the author.

Objective one: The author designed the study, set up the laboratory work, performed some of the laboratory assays, supervised some of the laboratory assays, performed the data analysis and interpretation.

Objective two: The author designed the study, performed all of the laboratory work, performed the data analysis and interpretation. The STATA code used in the selection of viral episodes to investigate from the parent study was written together with Patrick Munywoki, the principal investigator of the parent study. Some of the illustrative figures are by Patrick Munywoki, these have each been clearly cited.

Objective three: The author designed the study, set up the laboratory work, performed some of the laboratory assays, supervised some of the laboratory assays, including DNA extraction for microarray, but not the microarray itself, which was done in Jason Hinds'

laboratory at St George's University Hospital in London. The author performed the data analysis and interpretation.

Objective four: The author designed the study and set up the laboratory work in Kilifi. She drafted standard operating procedures for all the laboratories in the PERCH multi-site project. She supervised the laboratory work in Kilifi on Kilifi samples and performed the data analysis and interpretation for both the Kilifi sample set and the PERCH all-site sample set.

1.4.6 Publications during the PhD period

Morpeth SC, Huggett J, Murdoch DR, Scott JAG. Making standards for quantitative real-time pneumococcal PCR. Biomolecular Detection and Quantification. 2014. Accepted.

Hammitt LL, Ojal J, Bashraheil M, **Morpeth SC**, Karani A, Habib A, Borys D, Goldblatt D, Scott JA. Immunogenicity, impact on carriage and reactogenicity of 10-valent pneumococcal non-typeable *Haemophilus influenzae* protein D conjugate vaccine in Kenyan children aged 1-4 years: a randomized controlled trial. PLoS One. 2014 Jan 21;9(1):e85459

Hammitt LL, Kazungu **S**, **Morpeth** SC, Gibson DG, Mvera B, Brent AJ, Mwarumba **S**, Onyango CO, Bett A, Akech DO, Murdoch DR, Nokes DJ, Scott JA. A preliminary study of pneumonia etiology among hospitalized children in Kenya. Clin Infect Dis. 2012 Apr;54 Suppl 2:S190-9

Chapter Two

2 Background

Pneumonia is the leading cause of death in children <5 years of age globally, and in Africa.[1] *Streptococcus pneumoniae* is the leading cause of pneumonia,[2] causes invasive pneumococcal diseases (IPD) such as sepsis and meningitis, and also causes mucosal diseases such as otitis media and sinusitis. Pneumococcal disease is the leading cause of childhood vaccine-preventable death in the world.[11] Aetiological diagnosis of pneumonia for both epidemiological studies and clinical care is difficult; only ~13% of cases of radiologically-confirmed pneumonia are diagnosable by blood culture;[12] sputum is not spontaneously produced by young children and differentiating carriage from disease using culture-positivity of sputum or nasopharyngeal samples is difficult; diagnostic serology is insensitive in children and requires paired acute and convalescent sera;[13] and urinary antigen tests are not sufficiently specific for use in young children, presumably due to high rates of pneumococcal carriage.[14,15] Comprehensive pneumonia aetiology studies find high prevalence of putative viral and bacterial pathogens among cases and controls that are difficult to interpret [5] and interactions between respiratory viruses and pneumococci are likely to be important. Many of these diagnostic assays are easier to interpret for adults, but it is children under the age of five years, particularly in low and middle-income countries, who have the highest burden of disease and thus the greatest need for better diagnostic tests.

There are over 90 serotypes of pneumococcus that have been documented, with new types being discovered every year.[16,17] Prevention of IPD in children has largely been achieved in the industrialised world by means of the 7-, 10- or 13-valent pneumococcal

conjugate vaccines (PCV), providing protection against the most common invasive serotypes, but serotype-replacement disease is occurring and is being monitored.[18-21] In 2011 Kenya introduced into the childhood vaccination programme PCV-10, a 10-valent pneumococcal conjugate vaccine which covers 71% of the serotypes responsible for cases of IPD in children in East Africa.[22]

2.1 Pathogenesis of pneumococcal pneumonia and invasive pneumococcal disease

2.1.1 Biology of the pneumococcus

Streptococcus pneumoniae (the pneumococcus) is a gram-positive, encapsulated diplococcus that is an almost exclusive human pathogen. The ecological reservoir for the pneumococcus is the human nasopharynx, particularly the nasopharynges of young children, although adults can also be carriers.[23] IPD carries a high case-fatality, falling from 52-98% before antibiotics were available to 7-48% with the introduction of penicillin.[24] The 93+ serotypes of pneumococcus are based on the polysaccharide capsule, the main virulence determinant. Invasion and colonisation potential are based largely on capsule type, but within serotypes there are also inter-strain differences in invasion potential.[25] Indeed some common invasive serotypes, such as serotypes 1 and 5, are seldom found in carriage studies.[26,27]

A core genome has been proposed, comprising 73% of the pneumococcal genome, with some regions of diversity associated with invasion and others with colonisation.[28-30] The alternative coding DNA at the capsular locus alone consists of over 1.8Mbp, nearly as large as the entire pneumococcal genome.[16]

The natural state of pneumococcal colonisation is in biofilms covering the mucosal surfaces of the nasopharynx. The pneumococcus is capable of taking up extracellular DNA from other pneumococci and related species by natural transformation, known as competence, with an estimated 5% of the pneumococcal genome consisting of insertion sequences.[28] Entering the competent state is tightly regulated and involves quorum-sensing and fratricide.[31] The quorum-sensing mechanism may also produce bacteriocins, which can also affect non-competent pneumococci to lyse and release DNA that is also taken up for transformation.[32]

The pneumococcus exists in two phases, an opaque phase and a transparent phase, and can spontaneously switch back and forward between phases. The opaque phase has more capsular polysaccharide and less teichoic acid in the cell wall and the transparent phase expresses less capsular polysaccharide and more teichoic acid. In addition there are differences in expression of various cell surface proteins between phases. The transparent phase is better at adherence to host epithelium in the nasopharynx and lung, whereas the opaque phase is more virulent and better suited to invasion.[33,34] The polysaccharide capsule is crucial for evading opsonophagocytosis, protecting the cell-surface proteins from complement and antibodies, and as such is the most important pneumococcal virulence factor.

Pneumolysin is another key virulence factor. It is a pore-forming cytotoxin, lyses eukaryotic cells, and is also important for bacterial recognition by the innate immune system.[35] Pneumolysin is released from pneumococcus when the bacterial cells autolyse, but may also be exported from the cell walls of intact pneumococci.[36,37] Cell surface proteins associated with the pneumocccal cell wall such as pneumococcal surface adhesin A (PsaA), choline-binding proteins A and E (CbpA and CbpE), pilus protein, plasminogen

and fibronectin-binding proteins (PfbA and PfbB), pneumococcal adherence and virulence factors A and B (PavA and PavB) are important for adherence to nasopharyngeal epithelium.[36]

2.1.2 Biology of the interaction between pneumococcus and the host; from coloniser to invader

Certainly anti-capsular antibody provides protection against IPD in humans as is evidenced by the success of pneumococcal conjugate vaccination [18,20] but it is probably not the chief mechanism by which natural immunity abrogates the likelihood of IPD as unvaccinated children grow.[38] The host inflammatory response to pneumococcal growth and invasion is complex and requires modulation in order to control pneumococcus and limit host damage.[39]

Infants, the elderly and the immunocompromised are at most risk of IPD. The specific forms of immunocompromise that predispose to IPD are; B-cell defects (including either an intrinsic B-cell problem or impaired T-cell help), deficiencies of early components of the classical pathway of complement and C3 deficiency, asplenia, interleukin-1 receptor associated kinase-4 (IRAK-4) deficiency, myeloid differentiation factor (MyD88) deficiency,[40] impaired nuclear factor-kappaB (NF-KB) activation and HIV infection.[41-43] We can learn from these specific immunodeficiencies that antibody- and complement-mediated opsonization, splenic macrophages and the Toll-like receptor and interleukin-1 receptor pathway to inflammatory cytokine production are vital for protective immunity to pneumococcus.[39,40,42]

Transforming growth factor (TGF)-β1 and regulatory T-cells are key in tolerance of nasopharyngeal carriage of pneumococcus in mice and humans.[44] Clearance of

pneumococcal carriage by macrophages is an IL-17-secreting CD4+ T-cell (Th17)-dependent process.[35,45,46] Macrophages are important in controlling pneumococcal colonisation and also in preventing invasive and local pneumococcal disease.[47] Macrophage migration inhibitory factor (MIF)-deficient mice showed prolonged pneumococcal colonisation and had less macrophages in the upper respiratory tract than normal mice. Macrophages were dependent on pneumococcal pneumolysin to produce MIF. Mice deficient in the immune-modulating, small non-coding RNA, microRNA-155, were shown to have high-density colonisation and to be unable to clear pneumococcal carriage, with reduced macrophage recruitment to the nasopharynx, but remained able to prevent invasive disease.[45] Murine host gene expression in nasal lymphoid tissue in response to asymptomatic pneumococcal carriage began with a type 1 interferon inflammatory response for the first two weeks, followed by a T-cell response and mucosal epithelial healing.[48]

In a neonatal and infant mouse model, clearance of pneumococcal nasopharyngeal colonisation was delayed, infant mouse macrophages showed reduced cytokine production and IL-17A production from adult CD4+ T-cells was delayed when infant macrophages were used as antigen-presenting cells.[49] The reduction in risk for IPD after the first year of life in humans is thought to be at least partly due to maturation of the infant innate immune system.[38]

In a mouse model of immunosenescence, pneumococcal colonisation was prolonged in elderly compared to young adult mice, with delayed monocyte/macrophage influx and an altered innate immunity cytokine profile, but normal neutrophil responses. The authors postulated that impaired clearance of pneumococcal carriage could contribute to the increased risk of IPD in elderly humans.[50] In an adult human experimental challenge

model, pneumococcal carriage was shown to stimulate a mucosal and systemic antibody response that was protective against further carriage on repeat challenge, and passive transfer of these sera gave cross-serotype protection against invasive disease in mice.[51] The authors suggested that low carriage prevalence amongst elderly human populations might result in a lack of immune boosting and thus an increased risk of IPD.

Prior nasopharyngeal colonisation with pneumococcus is known to be protective against subsequent fulminant pneumonia and septicaemia in mice. The ability to prevent subsequent invasive disease in previously colonised mice was shown to be opsonophagocytic antibody-mediated protection from bacteraemia, as CD4+ T-cells were not required, despite induction of Th17 cytokines during re-challenge.[52] This is unlike the protection from subsequent colonisation afforded by prior nasopharyngeal colonisation with pneumococcus in the mouse, which is CD4+ T-cell dependent.[53] In humans, both antibody (anti-capsular polysaccharide and anti-protein) and Th17 responses are likely to be involved in protection from subsequent colonisation. [51]

Complement-deficiency is a well-known risk factor for IPD in humans.[42] Depleting mice of complement prior to challenge with intra-nasal pneumococcus resulted in the majority developing IPD, with the protection from complement being neutrophil and antibody-independent.[41] Both density of nasopharyngeal colonisation and density of infection of nasal lymphoid tissue were no different in complement-deficient or control mice prior to the development of septic symptoms, even though one-third of the complement-deficient mice were already bacteraemic. By contrast, density of nasopharyngeal colonisation among septic mice at the time of sacrifice was significantly higher than among non-septic mice. The authors concluded that complement does not prevent pneumococcal invasion of local lymphoid tissues but does prevent invasion into the

bloodstream. They suggested that colonisation might frequently result in pneumococcal bloodstream invasion, which is usually controlled by complement before IPD can become established. We do need to be careful in extrapolating directly from mouse models to human immunity. Studies have shown, for instance, that pneumococcal CbpA binds to factor H, an inhibitor of the complement alternative pathway, in humans, but does not bind to murine factor H.[54] In studies of human bacteraemic pneumonia, the most common invasive serotypes have not always been those with the highest case-fatality; it has been suggested that host factors such as mannose-binding lectin deficiency (part of the lectin pathway of complement activation) might allow less-invasive serotypes to invade and cause fatal disease.[55]

Pneumococci that do invade the bloodstream are subject to phagocytosis by neutrophils. Pneumococci deficient in nitric oxide (NO) defence systems were able to colonise mice but NO defence genes were needed for bloodstream invasion, presumably to cope with intracellular killing once phagocytosed by macrophages and neutrophils.[56] Pneumococci have been found to be capable of invading human erythrocytes in vitro and in the presence of erythrocytes were more able to evade neutrophil phagocytosis.[57] Opsonophagocytosis is enhanced by the immunising effect of prior colonisation. Mice previously colonised by pneumococcus were able to control and clear pneumococcal infection even when the bacteria were intravenously administered.[52]

In order to invade through the nasopharynx or, once in the bloodstream, into distal tissues such as the brain, pneumococci must be able to get through epithelial and endothelial cell barriers. CbpA, also known as pneumococcal surface protein C (PspC), interacts with the polymeric immunoglobulin receptor (pIgR) to achieve pneumococcal adherence and invasion into epithelial cells. pIgR exists to transport secretory IgA and IgM

across mucosal epithelium, and is then recycled back across the cell, during which pneumococci may take advantage of the system for ingestion and transcytosis across the epithelial cell barrier.[58] The platelet-activating factor receptor (PAFR) is implicated in adhesion, invasion and translocation of pneumococcus through endothelial cells, although it appears not to be an absolute requirement for the development of IPD.[59] Human endothelial cells have been shown to uptake pneumococci by both a clathrin and caveolae-dependent endocytosis in vitro; 90% of pneumococci were destroyed in the lysosome but 10% achieved transcytosis to the basal cell surface.[60]

Dispersion of pneumococci from nasopharyngeal biofilms could help explain the switch from asymptomatic carriage of pneumococci to the potential for invasion. Pneumococci dispersed from biofilms were shown to express virulence factors and were more likely to be in the opaque phase than pneumococci grown in planktonic culture or pneumococci from intact biofilms. Dispersion of pneumococci from biofilms in vitro and in vivo was achieved by infection with influenza A virus, or by application of host signals normally found in response to viral infection, such as hyperthermia, norepinephrine and extracytoplasmic ATP.[61] Transcriptome analysis demonstrated upregulation of carbohydrate metabolism and bacteriocins and downregulation of genes associated with colonisation (competence, fratricide and the transparent phenotype) among dispersed pneumococci.[62]

Pneumolysin is known to be an important virulence factor but we do not yet fully understand all of its effects beyond its cytotoxic capacity. Mice colonised with pneumolysin-deficient pneumococci had higher nasopharyngeal pneumococcal density than mice colonised with pneumolysin-expressing bacteria.[35] Pneumolysin activated an inflammatory response that led to pneumococcal clearance, a finding that has not always

been consistent. This could be because the effect of pneumolysin may depend on the genetic background or context of the pneumococcal strain in question.[63]

Pneumococcal nasopharyngeal colonisation, particularly with a high density of organisms, is thought to be a prerequisite for pneumonia and/or IPD, which is more likely to occur in the presence of viral upper respiratory tract infection.[64,65] It is believed that it is the recent nasopharyngeal acquisition of a virulent strain of an invasive serotype of *S. pneumoniae* that is the event that leads to disease if immune mechanisms fail. However, in the seminal study by Gray *et al* where this observation was made, 28 of the 31 pneumococcal infections observed were otitis media.[66,67]

2.1.3 Are respiratory viruses a 'trigger' for pneumococcal invasion?

Certainly both respiratory viruses and bacteria such as *Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis* and non-pneumococcal streptococci are all very commonly found in the nasopharynges of healthy children.[5,68] Respiratory viral infection in humans may result in upper respiratory tract symptoms such as coryza and nasal congestion. In young children, such symptoms may last at least 10 days.[69,70]

Temporal associations between circulating viral infections such as rhinovirus,[71] RSV and influenza [72,73] and IPD in children and adults have been documented in ecological studies. Influenza infection was associated with elevated blood pneumococcal concentration in IPD in South Africa.[74] Preceding influenza infection has been shown to be directly associated with severe pneumococcal pneumonia in children,[75] bacterial co-infection has been common among severe pandemic influenza infections [76] and considerable work has been done to investigate the relationship between respiratory viruses and pneumococcus in animal models.

2.1.3.1 In vitro studies

Mucosal inflammation in response to viral infection for many respiratory viruses increases host epithelial receptors, depending on the virus and on the cell type, augmenting pneumococcal binding and potentially promoting bacterial invasion.[67,77] For example, RSV, HPIV-3 and influenza virus all enhanced pneumococcal adhesion to human adenocarcinoma epithelial cells (A549) and to primary human bronchial and primary human small airway epithelial cells. RSV and HPIV-3 increased the expression of epithelial cell receptors for bacterial adherence by primary bronchial epithelial cells and A549 cells but not by primary small airway epithelial cells, and influenza virus did not increase receptor expression.[77] Human coronavirus NL63-infected human airway epithelial cells markedly increased adherence for pneumococcus at day 4 post viral infection, when viral replication was maximal.[78] Transcription of a set of host cell adhesion molecules was enhanced by infection of human pharyngeal cells by RSV, and adherence of pneumococci to the host cells was increased by a factor of 1.3- 2.0 in the presence of virus.[79] Platelet-activating factor (PAF) appears to be a receptor for adherence of *S. pneumoniae* to human respiratory epithelial cells, which is upregulated during RSV infection.[80] Additionally, RSV G-protein expressed on the surface of RSV-infected respiratory epithelial cells, was found to bind *S. pneumoniae*, further promoting bacterial adherence.[81,82] This increased bacterial adherence to human cell lines infected with RSV has been demonstrated for both nasopharyngeal cells and pneumocytes, and for a range of serotypes of *S. pneumoniae,* albeit with variation in degree of adherence between serotypes.[79,83]

As well as RSV, rhinovirus infection has been shown to increase adherence of *S. pneumoniae* to human tracheal epithelial cells via PAF.[84] Rhinovirus also increased

adherence of *S. pneumoniae* to human nasal epithelial cells, 1.5-fold, and increased expression of PAF from the cell lines.[85] Human coronavirus NL63 infection of human airway epithelial cells increased adherence of *S. pneumoniae*, but it was not clear how important the PAF receptor is in the process.[78]

Influenza virus infection does not seem to upregulate host epithelial cell receptors for bacterial adherence [77] but viral neuraminidase protein is probably important in promoting bacterial growth [78] and increases in free sialic acid due to influenza or other virus infection have been shown to augment pneumococcal nasopharyngeal biofilm density.[86] Depletion of host sialic acid availability limited influenza-induced pneumococcal replication.[87] Indeed, not only influenza virus infection, but also application of host responses to viral infection to epithelial cell lines or to mice, resulted in transformation of pneumococci from a biofilm or planktonic phenotype to a phenotype capable of virulent invasion.[61]

As well as increasing numbers of bacteria adhered to respiratory epithelium, respiratory virus infection has been shown to be associated with reduced clearance of bacteria from the respiratory tract. Human monocytes infected with RSV were less able to bind *S. pneumoniae*, showed reduced intracellular killing and reduced production of TNF-α.[88]

2.1.3.2 In vivo studies

Influenza A infections in infant mice already carrying nasopharyngeal pneumococcus increased pneumococcal concentration in the nasopharynx 50-fold, and in the lungs, 300-fold.[89] Transmission of pneumococcus between infant mice only occurred among those infected with influenza A. Influenza A reduced the numbers of pneumococcus required to establish pneumococcal pneumonia in mice from 100,000 to 100 colony-forming units

(CFU) and in turn pneumococcal infection increased established viral titres; influenza appeared to reduce the bacterial clearance ability of alveolar macrophages and pneumococcal infection enhanced viral release from infected cells.[90] Influenza virus not only increased pneumococcal nasopharyngeal density and carriage duration in mice, but influenza vaccination significantly reduced this excess pneumococcal carriage density.[91] Influenza virus infection in ferrets enhanced not only transmission of pneumococcal colonisation from donor ferrets, but considerably enhanced acquisition of pneumococcal colonisation in recipient ferrets, and also increased risk of disease, in a pneumococcal-strain dependent manner.[92] Increased density of pneumococcal colonisation in transmitting mice and lipopolysaccharide-induced nasopharyngeal inflammation but not neutrophil depletion in receiving mice also enhanced pneumococcal transmission in the mouse model.[93] In a mouse model of otitis media, influenza A co-infection overcame the difference in colonisation propensity and ability to cause otitis media between pneumococcal opaque or transparent phase variants.[94]

Infection with both RSV and pneumococcus in mice resulted in a higher incidence of pneumococcal bacteraemia and a higher bacteraemic pneumococcal concentration than in controls.[82] RSV infection reduced bacterial clearance and induced functional changes in the neutrophil response in mice. [95] Human metapneumovirus infection predisposed mice to severe pneumococcal pneumonia.[96]

2.1.3.3 In human children

Frequent carriage of pneumococcus in the first two years of life was associated with seroconversion to human metapneumovirus (hMPV) [97] but whether hMPV infection predisposed to pneumococcal carriage or the other way around is not clear. In a cross-sectional study of children attending day-care in Portugal, there was a significant

association between respiratory viral infection (78% of which was picornavirus infection, including rhinovirus) and pneumococcal colonisation density.[98] Among children with radiologically-confirmed pneumonia in Vietnam, nasopharyngeal pneumococcal concentration was 15-fold higher with viral co-infection than without viral co-infection and the difference was significant for influenza A, RSV and rhinoviruses. This finding was not observed among children with lower respiratory tract infection without radiologically-confirmed pneumonia.[99]

2.1.3.4 Summarising respiratory viral – pneumococcal interactions

We have direct evidence from animal studies and human cell lines that respiratory viral infections can increase pneumococcal nasopharyngeal adherence and decrease pneumococcal clearance, increasing pneumococcal concentration. We have ecological studies linking respiratory viral epidemics with pneumococcal disease, and we have evidence that severe pneumococcal pneumonia can follow influenza virus infection in pandemics. It may even be that respiratory viruses can trigger quiescent pneumococci colonising the nasopharynx in biofilms to change phase and become invasive. It has been postulated that respiratory viral infections may act synergistically with pneumococci of capsular or clonal types of lesser invasive potential, leading to IPD with such types.[100] Some have suggested that such viral-bacterial synergy may in fact account for all cases of IPD. [101]

2.1.3.5 Nasopharyngeal concentration and risk of disease

A high pneumococcal concentration in the upper airways may well be a risk factor for pneumonia or IPD but it could also be a result of pneumococcal pneumonia. The increased pulmonary secretions during pneumonia are expelled upwards by ciliary action

and either swallowed or expectorated. This could either seed the nasopharynx with pneumococci or provide direct sampling of such secretions during swab sampling.

Colonising pneumococci with invasive potential may have to reach a density threshold in the nasopharynx before they can breach the mucosa or spill over into the lungs in sufficient numbers to overwhelm the immune defences. Alternatively, the relationship between pneumococcal concentration in the nasopharynx and pneumococcal disease may be continuous. It may be oversimplifying the situation to consider all pneumococci to have the same relationship between nasopharyngeal concentration and likelihood of invasion. Respiratory viruses may trigger pneumococcal invasion - if so, do they do this by increasing pneumococcal carriage density? Young children are not only more likely to carry pneumococcus in the nasopharynx than older children or adults,[102] but they are also prone to carry a higher concentration of pneumococcus [103] and respiratory viruses are frequently present in this age group.[5] This high pneumococcal nasopharyngeal concentration could be part of the reason that young children are most at risk for pneumonia and invasive pneumococcal disease. A higher nasopharyngeal concentration of pneumococcal carriage could be part of the reason that children from lower socio-economic groups or under-resourced countries are at higher risk for pneumococcal pneumonia and IPD.

It may be that respiratory viral infection triggering an increase in pneumococcal nasopharyngeal concentration coinciding with recent acquisition of an invasive pneumococcal serotype is required for invasive disease to occur. However if density of pneumococcal carriage or triggering by respiratory viral infection is more important for the occurrence of invasion than recent acquisition of a virulent pneumococcal type, then duration of carriage increases in importance epidemiologically. Duration of carriage

declines with increasing age in young children, and with the number of carriage episodes.[66,104] Carriage duration [66] as well as other characteristics of carriage fitness,[105] and invasion potential,[34] is associated with serotype.

2.2 Molecular methods to diagnose pneumococcal pneumonia or invasive pneumococcal disease

2.2.1 Development of pneumococcal PCR assays

Molecular methods to diagnose pneumococcal pneumonia or IPD have been in development for the past three decades.[106-127] Some gene targets of nucleic acid amplification tests (NAATs) have been found to be non-specific to *S. pneumoniae*, present in non-pneumococcal *Streptococcus* species.[106,125,128] This is perhaps not surprising, given the propensity of *S. pneumoniae* to undergo genetic transformation, using donor DNA from other *Streptococcus* species such as *Streptococcus mitis* as well as other strains of pneumococcus.[129,130] A review of diverse studies carried out from 1993 to 2009 to assess PCR assays in the diagnosis of IPD from blood describes sensitivity of 57-66% and specificity of 88-99%, with specificity varying considerably depending on what sort of control groups were used.[124] In general, specificity was higher when the control group was healthy people than people with suspected but unproven IPD and was higher in adults than in children. Poor sensitivity has been thought to be due to low specimen volume that can be tested by NAATs, timing of sample collection compared to that for blood cultures, PCR-inhibitors and pneumococcal autolysis and DNA degradation from old specimens, specimens stored at -20°C, or subjected to freeze-thaw cycles.[116,122,125,127]

There have been PCR assays developed using the pneumolysin gene *ply*,[106,131,132] the surface adhesion protein *psaA*,[125] penicillin-binding protein PBP2b,[117] a gene of unknown

function spn9802,[122,133] the autolysin gene *lytA*,[116,125,134,135] capsular polysaccharide biosynthesis gene cpsA[136] and bacterial recombinase gene *recA*.[137] Dagan *et al* found that of 202 healthy controls, 35 (17%) were positive for *ply* PCR from serum samples. Carriers of pneumococci were more likely to be *ply* PCR positive from serum than non-carriers, and *ply*-positivity was highest among 2-year old children, the group with the highest nasopharyngeal pneumococcal carriage. Dagan and colleagues postulated that pneumococci or pneumococcal DNA may have been detectable in the bloodstream of some children that had escaped immune control in the nasopharynx but were cleared successfully without any clinical disease. We now know the pneumolysin gene as a PCR target detects not only pneumococcus but also other commensal streptococcal flora and Dagan *et al*'s suggestion has received little further attention as a result.[106,125]

2.2.2 *lytA* PCR

Although the autolysin gene has been demonstrated in some *S. pseudopneumoniae*[128] and *S. mitis* genomes,[108] the gene sequence can be distinguished from that in *S. pneumoniae*.[128,138] Most *lytA* assays have been demonstrated to have 100% analytical specificity for *S. pneumoniae*. At the CDC, a panel of 104 non-pneumococcal isolates including 11 non-pneumococcal viridans group streptococci and 10 *S. pseudopneumoniae* were negative by CDC-*lytA* PCR, although four of the *S. pseudopneumoniae* isolates were positive by the McAvin *lytA* PCR.[125,134] The *lytA* PCR published by Sheppard et al, was negative among all of 39 non-pneumococcal isolates including 18 *Streptococcus* species [116] and among all of 49 non-pneumococcal isolates including 22 mitis group streptococci.[138] The Messmer *lytA* PCR was negative among all of 51 viridans group streptococci, including 16 closely related streptococci that appeared pneumococcus-like when originally isolated.[120,139] The Nagai *lytA* PCR was negative among all of 108 non-

pneumococcal cocci, including 23 *S. pseudopneumoniae*, *S. mitis* or *S. oralis* strains and 54 other *Streptococcus* species.[136,140] A summary of these reports can be found in table 2.1 below. Worldwide experience with the *lytA* target for identification of IPD cases using real-time PCR from whole blood is increasing, with most groups now using the CDC-*lytA* assay.[74,121,141-147] It remains possible that the CDC-*lytA* assay will be demonstrated not to have 100% analytical specificity for pneumococcus, but there has been no evidence for this to date. Further testing against collections of closely-related, well-characterised non-pneumococcal streptococci would be helpful in this regard.

Table 2.1: Summary of published studies examining the specificity of *lytA* PCR for *S. pneumoniae* against closely related organisms

Reference	*lytA* PCR assay	No. of non-pneumococcal streptococci positive/tested	No. of *S. pseudopneumoniae* positive/tested
Carvalho *et al*, JCM 2007 (125)	CDC-*lytA*	0/11	0/10
Carvalho *et al*, JCM 2007 (125)	McAvin *lytA* (134)	0/11	4/10
Sheppard *et al*, JMM 2004 (116)	Sheppard *lytA*	0/18	
Greve & Moller, JMM 2012 (138)	Sheppard *lytA*	0/22	
Messmer *et al*, DMID 2004 (139)	Messmer *lytA*	0/51	
Park *et al*, JMM 2010 (136)	Nagai *lytA* (140)	0/54	0/23
Total isolates tested		**156**	**33**

2.2.3 Real-time *lytA* PCR from blood samples in the diagnosis of IPD

It appears that with current specimen handling and DNA extraction techniques, sensitivity of the *lytA* real-time PCR assay is now better than the previous gold standard for IPD, blood culture.[74,143,144] Specificity is more difficult to assess. In South Africa, *lytA* real-time PCR on fresh whole blood detected four times as many cases of IPD as blood cultures among adults and children, and only missed 1% of cases detected by blood culture, but there was no control group and no estimation of clinical specificity.[74] In Australia, a *lytA* real-time PCR performed on 1,800 adults presenting to the emergency department

alongside standard blood cultures detected 11 of 11 blood-culture-positive cases of IPD and 9 blood-culture-negative cases which may have been IPD (clinical scenario compatible with IPD and no other pathogen isolated).[141]

In Italy *lytA* real-time PCR from blood was negative among 147 well children, half of whom were demonstrably carrying pneumococcus in the nasopharynx [15] and in South Africa both *lytA* and *psaA* real-time PCR's from serum were negative in 100 well children with a mean age of 5.8 years, half of whom were carrying pneumococcus in the nasopharynx.[148] In Sweden, *lytA* PCR was positive in plasma of 9% of 92 adults with community-acquired pneumonia and none (0%) of 91 adult control patients hospitalised for skin or urinary tract infections, or arthritis or elective surgery, without respiratory symptoms.[122]

The lack of a sensitive gold-standard case-definition for IPD, and the use of well children or adults who are not the target population for the diagnostic assay, in these types of studies, produces a biased estimate of clinical specificity. A more appropriate control group would be children admitted to hospital with possible IPD who are proven not to have IPD; such a control group is almost impossible to find because we know that blood cultures themselves are insensitive in the diagnosis of IPD.

In Slovenia, *lytA* PCR was positive in plasma among 32% of 160 children and 23% of 180 adults with community-acquired pneumonia.[146] The authors also ran the assay on controls; *lytA* PCR was positive in plasma among none (0%) of 50 adults being investigated for Lyme disease and two (7%) of 29 children with acute febrile illness that was not pneumonia and in whom non-pneumococcal infections had been identified. The mean age of the childhood controls was 1.5 years, and both control children that were *lytA* PCR positive from plasma carried pneumococcus in the nasopharynx.

2.2.4 Quantitative real-time *lytA* PCR from blood samples in IPD

Use of quantitative PCR from blood samples as a marker of pneumococcal disease severity is also receiving interest, with higher pneumococcal concentrations associated with mortality. [74,145,147,149] Quantitative real-time *lytA* PCR from blood might also assist in separating children with bacteraemia from children with PCR-positivity in blood ("DNAemia") from high nasopharyngeal concentration, or escape of pneumococcus into the bloodstream without resulting sepsis, if such a phenomenon really occurs. It is possible that pneumococcal DNA could be detectable in blood specimens more often than we realise. Pneumococcal invasion into the bloodstream may occur more frequently than IPD as diagnosed by blood cultures from a clinically ill child occurs, but be controlled by complement-mediated bacterial lysis and opsonophagocytosis. The frequency with which such a phenomenon might occur would depend on the population carriage prevalence and density of carriage, the age structure of the population, the attack rate of circulating respiratory viruses, exposure to smoke and pollution, host genetics and the circulating bacterial strains.

2.2.5 Nasopharyngeal pneumococcal concentration in the diagnosis of pneumococcal pneumonia

A high nasopharyngeal concentration of pneumococcus has been associated with pneumonia in HIV-infected adults in South Africa [126] and with radiologically-confirmed pneumonia among children in Vietnam [99] and Italy.[150] In South Africa, high nasopharyngeal pneumococcal concentration has been associated with severity of pneumococcal pneumonia in HIV-infected adults,[151] and with IPD in adults and children with pneumonia.[152]

Young children do not expectorate sputum so lower respiratory tract specimens in this age group are not easy to collect. Nasopharyngeal swabs, washes or aspirates can be used to sample the upper respiratory tract and induced sputum or bronchoalveolar lavage are uncommonly performed to access the lower respiratory tract. Detection of *S. pneumoniae* in respiratory tract specimens of infants and young children by culture or PCR is very common, as seen in table 2.2; understanding the clinical significance of it in the diagnosis of pneumonia is difficult.[5,115,146,150,153,154]

Table 2.2: Detection of *S. pneumoniae* in paediatric respiratory tract specimens in the diagnosis of pneumonia. Detection is common, but understanding the clinical significance of it in the diagnosis of pneumonia is not straightforward.

Reference	Specimen type	Method	No. of samples positive for *S. pneumoniae*	No. of samples tested	Percentage positive
Hammitt *et al*, CID 2012 (5)	Induced sputum from children	Culture, only predominant potential pathogens	16	417	4%
Morozumi *et al*, JCM 2006 (115)	375 Nasopharyngeal secretions & 14 throat swabs from children, 40 sputum samples from adults	Real-time multiplex PCR and culture	202	429	47%
Cvitkovic Spik *et al*, Scand JID 2013 (146)	Nasopharyngeal swabs from children	Real-time PCR	141	160	88%
Esposito *et al*, PIDJ 2013 (150)	Nasopharyngeal swabs from children	Real-time PCR of samples positive for RSV or rhinovirus	148	260	57%
Hamano-Hasegawa *et al*, JIC 2008 (153)	Nasopharyngeal swabs from children	Real-time PCR	415	1700	24%
Lahti *et al*, Thorax 2009 (154)	Induced sputum from children	Culture of good quality samples	35	76	46%

There may be a continuum between low-level asymptomatic carriage of pneumococcus in the nasopharynges of healthy children, to increased carriage during periods of immune-stress or mucosal damage such as during exposure to smoke or viral infection, to local mucosal disease such as sinusitis and otitis media, to highest nasopharyngeal concentrations associated with pneumonia or IPD. If nasopharyngeal pneumococcal concentration by quantitative real-time PCR is greater among children with pneumococcal pneumonia than among children with viral respiratory tract infections or well children then it might be possible to look for a threshold correlating with pneumococcal pneumonia, as has been shown for adults.[126,155] It has been assumed in the adult study in South Africa, that the increased nasopharyngeal concentration was in the causal pathway for pneumococcal pneumonia, which might be the case, but in fact we don't know the direction of causality for this association. In Vietnam the median pneumococcal concentration in the nasopharynx was 7.8×10^6/mL in children under the age of five years with radiologically-confirmed pneumonia compared to 1.3×10^6/mL in children with other lower respiratory tract infection, or 7.9×10^5/mL in healthy children ($p<0.001$ for each comparison), but there was no suitable threshold for potential use as a diagnostic assay for pneumococcal pneumonia.[99]

The ideal study to describe the dynamics of the spread of pneumococci in the body from healthy carriage through to bacteraemic pneumonia would be a cohort study of healthy children with frequent sampling of the nasopharynx and blood, from the asymptomatic state, through mild upper respiratory tract illness and into the development of naturally occurring pneumonia and pneumococcal bacteraemia. Unfortunately such a cohort would have to be extremely large in order to observe a rare event such as IPD in sufficient numbers, and the frequency of sampling needed would not be tolerated. A more epidemiologically efficient solution would be a series of cross-sectional samples of

suitable populations of children along the continuum outlined above, to describe the pneumococcal concentration in the nasopharynx and how this relates to pneumococci detectable in the blood.

2.3 Nasopharyngeal pneumococcal concentration and pneumococcal conjugate vaccines

As previously noted, there are more than 90 serotypes of pneumococcus, and only some of them have the propensity to cause invasive disease. Pneumococcal conjugate vaccines have been shown to reduce nasopharyngeal carriage of vaccine serotypes,[102,156-159] including in people not themselves vaccinated.[160,161] Reduction in carriage among vaccinated individuals and less transmission to unvaccinated individuals leads to less invasive disease among the unvaccinated as well as the vaccinated. These are indirect effects of vaccination – also known as the herd effect. The herd effect has played a large part in overall IPD reduction.[18] Pneumococcal conjugate vaccination is thought to bring about less carriage of vaccine-types by inducing mucosal immunity but exactly how this works is unclear.[64] IgG and IgA antibodies to capsular polysaccharide may be present at the mucosal surface in response to vaccination, in line with the increase in systemic IgG antibody production, but in the mouse model antibodies do not appear to be necessary for clearance of pneumococcal carriage.[162] The decrease in carriage of vaccine-types associated with PCV vaccination may decay with time over subsequent years so that mucosal immunity is lost, while immunity to invasive disease is retained.[163]

Reduction in IPD caused by vaccine-types has been concurrent with an increase in IPD caused by non-vaccine-types, although in most populations this has not been sufficient to

negate the overall impact of vaccination on IPD.[3,18,20] Serotype replacement disease would be particularly a problem if more antibiotic resistant strains replace the vaccine-types in circulation. The drop in number of IPD cases in response to vaccination means that monitoring the impact of pneumococcal vaccines, whether the currently available polysaccharide-protein conjugated vaccines or new generation protein or whole cell vaccines, benefits from the use of carriage studies in addition to IPD surveillance; carriage is so much more common than IPD that it is easier to study.[164] It is possible to separate out the direct and indirect effects of vaccination using carriage studies as part of vaccine impact monitoring. The vaccine effect on carriage in an individual is a direct marker of protection against disease (protection against acquisition, density or duration of carriage and protection against progression to disease).[64] The indirect effect of vaccination (the herd effect) depends on the direct effect on carriage in order to reduce transmission.[164] It would therefore be useful to examine the impact of vaccination on concentration of vaccine-type (VT) carriage, not only on VT carriage prevalence.

2.3.1 Detecting multiple serotype carriage for vaccine studies

Children can carry more than one serotype of pneumococcus at a time, and do so in up to 50% of children sampled using current methods,[19,165-168] perhaps even up to 67% as measured by an Italian group – although this study used nasopharyngeal swabs collected via the oropharynx and a real-time PCR method that could potentially detect capsular genes among non-pneumococcal streptococci in addition to those of pneumococcus.[163]

Historically, multiple serotype carriage was detected by injecting aliquots of saliva intraperitoneally into mice. Specimens yielding a pneumococcus were subsequently injected into mice pretreated with antiserum to the serotype initially recovered and the procedure repeated until no further serotypes were recovered.[23] Using this method,

Austrian reports that Gundel and Okura found in 1933 that 70 of 95 school children were harbouring more than one pneumococcal serotype.[23] Subsequently, culture of nasopharyngeal swabs on 5% blood agar media, with or without gentamicin, was favoured as a more practical method for large-scale studies of nasopharyngeal carriage. Gentamicin was found to increase the sensitivity of the blood agar plate method for throat swabs in the detection of pneumococcus.[169] Serotyping more than one colony of pneumococcus from culture plates was found to be both labour-intensive and inadequate to detect multiple serotype carriage because of the dominance of the prevailing type.[170] This method was standardised by a WHO working group but it was recognised that multiple serotype carriage was not readily detected and a call for improved methods was made.[168]

The PneuCarriage project (http://www.thl.fi/roko/pneumocarr/pneucarriage_project.html), an international consortium funded by the Gates Foundation, recently set out to establish the optimum methods for detection of multiple serotype carriage in order to publish updated recommended standard methods.[171] Such standardisation is needed if we are to be able to measure and compare pneumococcal vaccine effectiveness in different settings around the world. Fifteen different methods were studied using a standard set of 81 spiked simulated samples, the best five methods being; pneumococcal microarray,[166] latex agglutination of a sweep from the culture plate,[166] restriction fragment length polymorphism (RFLP) analysis,[172] real-time PCR from cultured isolates [142] and real-time PCR directly from nasopharyngeal specimens.[142,163] These five methods went forward to test a set of 260 nasopharyngeal swab samples and in addition traditional serotyping of 100 colonies was performed for each sample. The gold standard was a serotype found by at least two different methods. Among these five methods the best scientific

performance was displayed by the pneumococcal microarray at St George's University Hospital in London (http://www.bugs.sgul.ac.uk/node/1), with 96% sensitivity and 94% positive predictive value.[171]

2.3.2 Assessing pneumococcal carriage concentration in vaccine studies

Carriage density of vaccine-types, assessed by semi-quantitative culture, decreased in the vaccine-arm of a PCV-7 randomised controlled trial among a Native American population.[158] In a cluster-randomised trial in the Gambia, carriage density was also measured by semi-quantitative culture and found to decrease for both vaccine and non-vaccine types after administration of PCV-7.[103] Antibodies to polysaccharide capsule were higher among individuals from fully-vaccinated than from control (infant vaccination only) villages, and increased with age to a plateau at 10 years of age.[173] Pneumococcal carriage density was highest in the youngest children and decreased with age, beyond the plateau in polysaccharide antibody concentration, suggesting such antibodies were not solely responsible for the changes in carriage.[103] In Kenya, vaccine-type carriage prevalence decreased after PCV-10 was given to toddlers in a randomised controlled trial, however concentration of vaccine-type carriage was not assessed.[10] Density of nasopharyngeal colonisation by pneumococcus is presumably important for both transmission and invasive disease [64] but we don't know what the nature of the relationships are.

Carriage prevalence as detected by culture is likely to be an indicator of colonisation density, but the relationship between culture positivity and colonisation density by qPCR seems to vary by method. In an experimental human pneumococcal carriage model, culture of fresh nasal washes from adults was able to detect down to 10^1 CFU/mL.[174,175] In Peruvian children sampled by rayon nasopharyngeal swab into skim-milk tryptone glucose glycerol (STGG) transport media, stored at -70°C, culture was significantly more likely to

detect pneumococcus at a density of >10^5 CFU/mL by qPCR.[176] By measuring the concentration of pneumococcus in a nasopharyngeal sample, whether by a sensitive quantitative culture method or by quantitative PCR, we assume we will get more information about changes in carriage associated with vaccination or viral infection.

A *lytA* qPCR will only tell us the overall concentration of pneumococcus in the nasopharynx since the autolysin gene is common to all pneumococci and is not serotype-specific. It would be worth knowing, not only what the concentration of nasopharyngeal colonisation is, but also what the concentration of colonisation with vaccine and non-vaccine types, and how that changes after vaccination. This is pertinent because we expect a direct vaccine effect on vaccine serotypes only, and because of the possibility of serotype replacement disease; IPD caused by non-vaccine serotypes. This question could be answered with the pneumococcal microarray revealing the relative proportions of each serotype present, together with the *lytA* qPCR. It would also be possible to use a singleplex quantitative PCR for each serotype that might be present, as long as the same standards could be used to generate each standard curve in the PCR assays. Potential difficulties with this latter approach include the fact that capsular gene targets for serotyping assays may be found in non-pneumococcal streptococci that are also present in nasopharyngeal or oropharyngeal swab samples.[177,178]

Chapter Three

3 Initial *lytA* real-time PCR method validation

3.1 Introduction

The *lytA* quantitative real-time PCR (*lytA* qPCR) method used in this work was originally developed at the Centers for Disease Control and Prevention (CDC) in Atlanta [125] and was developed further by Dr Nicole Wolter at the National Institute for Communicable Diseases (NICD) Laboratory in Johannesburg, South Africa. In brief, the method used DNA extracted from a standard volume whole blood in a singleplex real-time PCR reaction run alongside a set of standards made from DNA extracts of pneumococcal isolates run in triplicate in each plate. From the resulting standard curve, the Cycle Threshold (Ct) value of the PCR reaction could be converted into a quantity in copies/mL.

The *lytA* qPCR assay in use at NICD was transferred to Kilifi and initial method validation was performed to ensure the method had analytical sensitivity and specificity for pneumococci found in Kilifi. Methods for making quantitation standards for the assay were explored and experiments were performed to choose an appropriate matrix for dilution of qPCR standards. The effect of different template volumes for the assay was investigated. The lower limit of detection and lower limit of linearity of the assay were examined.

The final methods used across the PhD project for whole blood and nasopharyngeal samples can be seen in chapter seven: General Laboratory Methods.

3.2 Analytical sensitivity and specificity

As discussed in detail the chapter two, the *lytA* target is said to have 100% specificity for *S. pneumoniae*. Having not used the assay in Kilifi before, we first aimed to demonstrate analytical sensitivity and specificity on local isolates.

3.2.1 Methods

The *lytA* PCR method as described in section 7.11 in chapter seven was performed on; the type strain of pneumococcus, ATCC 49619, 30 local strains of pneumococcus that had caused invasive pneumococcal disease among children in Kilifi, 30 strains of non-pneumococcal streptococci from the same collection, and 70 strains of pneumococcus from nasopharyngeal carriage studies in Kilifi. DNA extracts for these isolates were a 1:10 dilution of samples extracted by a simple boilate method, (described in section 7.8, chapter seven) which we were using in our laboratory for DNA extraction of invasive isolates for the conventional pneumococcal 'serotyping' PCR, adapted from the CDC method.[179]

3.2.2 Results

The type strain of pneumococcus, the 30 local pneumococcal isolates that had caused IPD and the 70 carriage isolates, including both serotypable and non-serotypable strains, were all successfully amplified by the *lytA* PCR assay. The 30 non-pneumococcal streptococci were not amplified by the assay, giving an analytical sensitivity and specificity of 100%, from isolates. The Ct values of these isolates can be seen in Table 3.1.

Table 3.1: Analytical sensitivity and specificity of *lytA* real-time PCR for *Streptococcus pneumoniae* in Kilifi

Isolate no.	Genus	Species/Group	Serotype	*lytA* Ct value
36003	*Streptococcus*	viridans group		negative
36191	*Streptococcus*	Group F		negative
36199	*Streptococcus*	Group D		negative
36205	*Streptococcus*	Group B		negative
36365	*Streptococcus*	species		negative
36679	*Streptococcus*	viridans group		negative
36780	*Streptococcus*	viridans group		negative
38044	*Streptococcus*	Group F		negative
38436	*Streptococcus*	Group D		negative
38451	*Streptococcus*	viridans group		negative
38461	*Streptococcus*	Group D		negative
38526	*Streptococcus*	Group D		negative
38527	*Streptococcus*	Group F		negative
38585	*Streptococcus*	Group D		negative
38783	*Streptococcus*	Group B		negative
38794	*Streptococcus*	Group D		negative
38795	*Streptococcus*	viridans group		negative
38861	*Streptococcus*	Group D		negative
39247	*Streptococcus*	Group B		negative
39258	*Streptococcus*	Group B		negative
39281	*Streptococcus*	Group D		negative
39421	*Streptococcus*	viridans group		negative
40348	*Streptococcus*	viridans group		negative
40818	*Streptococcus*	Group A		negative
41274	*Streptococcus*	viridans group		negative
42122	*Streptococcus*	viridans group		negative
42250	*Streptococcus*	Group G		negative
42771	*Streptococcus*	Group A		negative
44867	*Streptococcus*	viridans group		negative
44871	*Streptococcus*	viridans group		negative
44929	*Streptococcus*	viridans group		negative
41488	*Streptococcus*	*pneumoniae*	23F	11.50
43188	*Streptococcus*	*pneumoniae*	18C	11.85
45165	*Streptococcus*	*pneumoniae*	1	11.93
41078	*Streptococcus*	*pneumoniae*	23A	11.96
43874	*Streptococcus*	*pneumoniae*	5	12.14
44608	*Streptococcus*	*pneumoniae*	5	12.36
42860	*Streptococcus*	*pneumoniae*	6A	12.45
43969	*Streptococcus*	*pneumoniae*	1	12.48
43050	*Streptococcus*	*pneumoniae*	6B	12.53
44261	*Streptococcus*	*pneumoniae*	1	12.61
44656	*Streptococcus*	*pneumoniae*	14	12.68
44655	*Streptococcus*	*pneumoniae*	1	12.84
44527	*Streptococcus*	*pneumoniae*	14	12.85
43639	*Streptococcus*	*pneumoniae*	12F	13.19
44371	*Streptococcus*	*pneumoniae*	14	13.27
44781	*Streptococcus*	*pneumoniae*	13	13.40

Isolate no.	Genus	Species/Group	Serotype	*lytA* Ct value
44346	*Streptococcus*	*pneumoniae*	6B	13.62
43429	*Streptococcus*	*pneumoniae*	19A	13.83
44516	*Streptococcus*	*pneumoniae*	2	13.83
43203	*Streptococcus*	*pneumoniae*	6A	13.89
44746	*Streptococcus*	*pneumoniae*	19F	14.23
43766	*Streptococcus*	*pneumoniae*	3	14.86
38595	*Streptococcus*	*pneumoniae*	15A	18.18
36880	*Streptococcus*	*pneumoniae*	10A	18.28
37423	*Streptococcus*	*pneumoniae*	4	18.30
39316	*Streptococcus*	*pneumoniae*	1	18.55
36827	*Streptococcus*	*pneumoniae*	11A	18.77
38827	*Streptococcus*	*pneumoniae*	9V	19.04
38379	*Streptococcus*	*pneumoniae*	5	19.09
40169	*Streptococcus*	*pneumoniae*	15B	19.18
38039	*Streptococcus*	*pneumoniae*	13	19.30
37013	*Streptococcus*	*pneumoniae*	21	37.15
2045	*Streptococcus*	*pneumoniae*	untyped	15.92
3113	*Streptococcus*	*pneumoniae*	untyped	17.49
3201	*Streptococcus*	*pneumoniae*	untyped	18.87
3391	*Streptococcus*	*pneumoniae*	untyped	18.44
30581	*Streptococcus*	*pneumoniae*	nontypable	15.64
30831	*Streptococcus*	*pneumoniae*	nontypable	17.69
31473	*Streptococcus*	*pneumoniae*	nontypable	18.19
31477	*Streptococcus*	*pneumoniae*	nontypable	16.96
31516	*Streptococcus*	*pneumoniae*	nontypable	17.36
32303	*Streptococcus*	*pneumoniae*	nontypable	18.91
34531	*Streptococcus*	*pneumoniae*	nontypable	17.54
34749	*Streptococcus*	*pneumoniae*	nontypable	18.34
34941	*Streptococcus*	*pneumoniae*	nontypable	17.72
34991	*Streptococcus*	*pneumoniae*	nontypable	17.17
35279	*Streptococcus*	*pneumoniae*	nontypable	17.62
35328	*Streptococcus*	*pneumoniae*	nontypable	17.84
35396	*Streptococcus*	*pneumoniae*	nontypable	18.00
35421	*Streptococcus*	*pneumoniae*	nontypable	17.82
35437	*Streptococcus*	*pneumoniae*	nontypable	17.18
44229	*Streptococcus*	*pneumoniae*	19F	16.40
44231	*Streptococcus*	*pneumoniae*	8	17.51
44280	*Streptococcus*	*pneumoniae*	9N	15.78
44282	*Streptococcus*	*pneumoniae*	6A	18.69
44285	*Streptococcus*	*pneumoniae*	23F	16.55
44287	*Streptococcus*	*pneumoniae*	6B	17.75
44291	*Streptococcus*	*pneumoniae*	3	16.96
44292	*Streptococcus*	*pneumoniae*	13	19.59
44395	*Streptococcus*	*pneumoniae*	24F	15.45
44406	*Streptococcus*	*pneumoniae*	15A	18.42
44408	*Streptococcus*	*pneumoniae*	18F	17.23
44411	*Streptococcus*	*pneumoniae*	6A	17.38
44470	*Streptococcus*	*pneumoniae*	18A	14.95
44503	*Streptococcus*	*pneumoniae*	23F	16.70
44519	*Streptococcus*	*pneumoniae*	34	17.20
44615	*Streptococcus*	*pneumoniae*	21	16.77

Isolate no.	Genus	Species/Group	Serotype	*lytA* Ct value
44621	*Streptococcus*	*pneumoniae*	23B	17.30
44624	*Streptococcus*	*pneumoniae*	6A	18.98
44672	*Streptococcus*	*pneumoniae*	15C	16.35
44679	*Streptococcus*	*pneumoniae*	14	18.23
44691	*Streptococcus*	*pneumoniae*	15A	16.38
44692	*Streptococcus*	*pneumoniae*	6B	17.29
44695	*Streptococcus*	*pneumoniae*	19F	17.54
44716	*Streptococcus*	*pneumoniae*	13	16.72
44808	*Streptococcus*	*pneumoniae*	11A	16.53
44810	*Streptococcus*	*pneumoniae*	5	16.86
44814	*Streptococcus*	*pneumoniae*	22F	17.59
44818	*Streptococcus*	*pneumoniae*	35B	18.16
44819	*Streptococcus*	*pneumoniae*	1	16.78
44820	*Streptococcus*	*pneumoniae*	16F	17.55
44825	*Streptococcus*	*pneumoniae*	3	19.58
44830	*Streptococcus*	*pneumoniae*	7F	17.95
44839	*Streptococcus*	*pneumoniae*	20	17.13
44910	*Streptococcus*	*pneumoniae*	11A	15.79
44914	*Streptococcus*	*pneumoniae*	19C	15.21
44987	*Streptococcus*	*pneumoniae*	35F	18.32
45013	*Streptococcus*	*pneumoniae*	6B	18.89
45051	*Streptococcus*	*pneumoniae*	6A	18.95
45052	*Streptococcus*	*pneumoniae*	15B	16.99
45057	*Streptococcus*	*pneumoniae*	18C	15.29
45061	*Streptococcus*	*pneumoniae*	19A	16.65
45062	*Streptococcus*	*pneumoniae*	35A	16.25
45079	*Streptococcus*	*pneumoniae*	23F	17.55
45084	*Streptococcus*	*pneumoniae*	6B	16.25
45128	*Streptococcus*	*pneumoniae*	19F	16.91
45131	*Streptococcus*	*pneumoniae*	9V	17.59
45227	*Streptococcus*	*pneumoniae*	13	16.41
45237	*Streptococcus*	*pneumoniae*	10A	15.87
45239	*Streptococcus*	*pneumoniae*	6A	17.39
45297	*Streptococcus*	*pneumoniae*	19F	18.45
90617	*Streptococcus*	*pneumoniae*	untyped	17.32

3.2.3 Conclusions

This initial validation demonstrated that the CDC primers and probe were able to detect a range of locally circulating pneumococci in Kilifi with 100% analytical sensitivity and specificity. The viridans group streptococci were alpha-haemolytic streptococci from blood cultures at Kilifi County Hospital that were not *S. pneumoniae* because they were resistant to optochin, but we were not able to characterize them further. It would have

been preferable to have well-characterized local isolates of *S. pseudopneumoniae* and *S. oralis* with which to test analytical specificity, but this has been done at other centres, as discussed in chapter two.

3.3 Making standards for quantitative *lytA* PCR

The work described in this section has been accepted for publication.

3.3.1 Introduction

We set out to explore methods for making in-house standards for the quantitative real-time *lytA* PCR assay. Research and Clinical Microbiology laboratories alike are embracing qPCR for detection of putative microbial pathogens where simply presence of a target organism is insufficient to determine pathogenicity.[99,126,147,180,181] For example, *Pneumocystis jirovecii* is known to cause severe pneumonia, but upon the advent of real-time PCR diagnostics, it was discovered that sensitive qualitative PCR is generally non-specific. It is frequently positive in well patients or those with another cause of illness. Quantitative PCR, however, may predict which samples are indicative of *P. jirovecii* pneumonia.[181] Likewise, detection of *Streptococcus pneumoniae* from respiratory tract samples by quantitative real-time PCR may distinguish nasopharyngeal carriage of commensal *S. pneumoniae* from pneumococcal pneumonia.[126] There is increasing interest globally in quantitative real-time PCR for detection of pneumococcus from respiratory tract samples, particularly utilizing the *lytA* assay. [125,126]

qPCR requires the use of standards of known concentration, serially diluted to form a linear relationship between the quantification cycle threshold (Ct) value and the

logarithmic value of the standard concentration. Methods for making these standards are not always described in publications.[182]

3.3.2 Methods

One possible method was to make a suspension of pneumococcus, plate out serial dilutions of the suspension, perform colony counting, and calculate the concentration of the original suspension in colony-forming-units (CFU)/mL from which DNA would be extracted for the standards.[149] Another method was to extract the DNA from a suspension of pneumococcus, measure the DNA concentration in ηg/μL, from which the concentration in genome-copies/mL calculated based on molarity (using Avogadro's number) (personal communication, Nicole Wolter, NICD). These methods are unlikely to be equivalent because they assume a single viable bacterial cell produces a single colony-forming unit; may be perfectly extracted by DNA extraction methods and will contain only one genome, with only one gene copy. [183] If they were the same then the ratio of the concentration in CFU/mL to the concentration in genome-copies/mL would be one. To investigate the impact of pneumococcal autolysis, we used *Escherichia coli* as a comparator organism. For the same reason, we used bacterial growth in broth at log phase as well as bacterial growth harvested from solid media, to make the starting-point suspensions.

S. pneumoniae ATCC 49619 and *E. coli* ATCC 25922 were grown to log phase in brain heart infusion broth, or harvested from an overnight culture plate on blood agar and suspended in normal saline, and adjusted to 3.0 McFarland. The suspensions were serially diluted 1:5 nine times and 100μL of each of the last three dilutions plated in duplicate on blood agar for incubation overnight at 35+/- 2°C in 5% CO_2. In addition, 1mL of the suspension was used for immediate extraction. To estimate CFU/mL, all countable plates underwent

colony counting on the following day to calculate the concentration of the original suspension.

DNA was extracted (QIAamp DNA mini kit, Qiagen, Germany) with two elution steps as described by the manufacturer to optionally maximise DNA recovery, and DNA concentration estimated in ηg/μL using the Nanodrop™ spectrophotometer (ThermoScientific, USA). Genome copies were estimated using the formula mass = DNA size (base pairs) x 1mole/6.023e23 molecules x 660g/mole. Twelve experiments were performed per organism, by the same operator, on different days.

The ratio of quantities derived from colony counting/DNA concentration (CFU/genome-copies) were compared by suspension method and by organism, using the Wilcoxon rank sum test.

3.3.3 Results

The distribution of bacterial concentrations by colony counting (mean log [sd] concentration for *S. pneumoniae* 8.4 [0.7] CFU/mL and for *E.coli* 8.9 [0.4] CFU/mL) was wider than the distribution of bacterial concentrations as calculated by DNA concentration (mean log concentration for *S. pneumoniae* 8.9 [0.2] genome-copies/mL and for *E. coli* 8.6 [0.2] genome-copies/mL). The distributions can be seen in Figure 3.1. For *S. pneumoniae* Spearman's rank correlation between the bacterial concentration as measured by colony counting and the bacterial concentration as calculated from DNA concentration was 0.69 (p=0.014) and for *E. coli* this was 0.10 (p=0.746). The correlations can be seen in figure 3.2. The ratio of quantities derived from the two methods was almost never one, and can be seen in Figure 3.3.

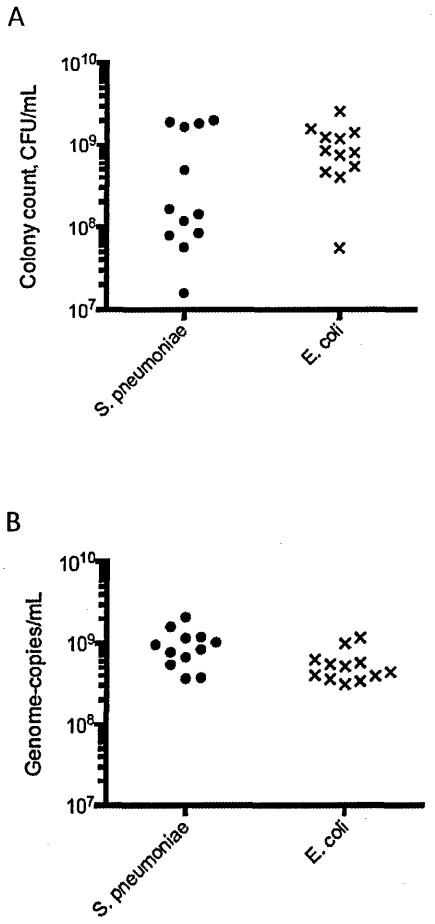


Figure 3.1: Spread of bacterial concentrations of standard suspensions of *S. pneumoniae* and *E. coli.* A: Obtained by colony counting in CFU/mL. **B:** Obtained by DNA concentration in genome-copies/mL.

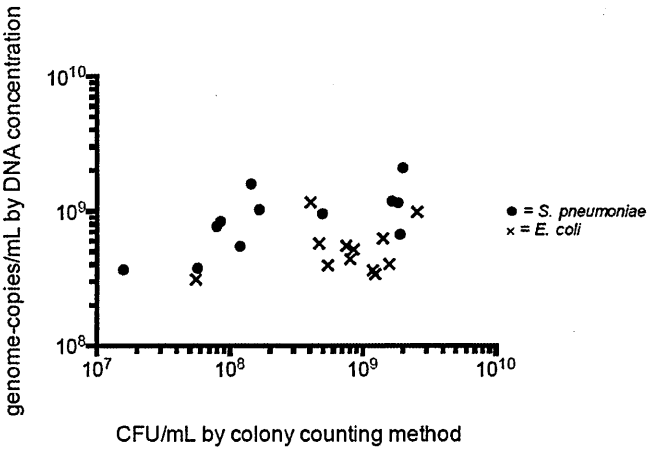


Figure 3.2: Poor correlation between methods of measuring bacterial concentration. Colony counting in CFU/mL and DNA concentration in genome-copies/mL.

The median (IQR) ratio (CFU/genome-copies) for *S. pneumoniae* combining both culture methods was 0.19 (0.1-1.2) and for *E. coli* was 1.74 (1.1-2.9), p=0.007. For either *S. pneumoniae* or *E. coli*, there was no difference in the ratio (CFU/genome-copies) whether culture on solid media or log phase growth in broth was used to make the suspension, as can be seen in figure 3.3.

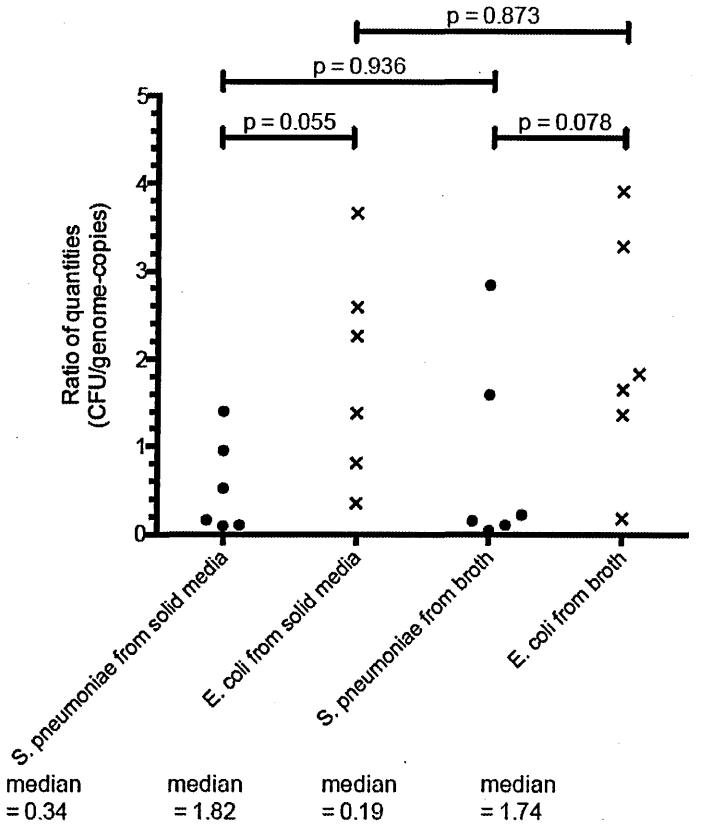


Figure 3.3: Ratio of quantification methods. Colony counting/DNA concentration in CFU/genome-copies, by organism and culture-medium.

3.3.4 Discussion

There was poor correlation between the bacterial concentration as measured by colony counting and the bacterial concentration as calculated from DNA concentration.

The ratio of bacterial concentration by colony counting to the bacterial concentration as calculated from DNA concentration, whether the measured suspension was made in saline from growth on solid media or from broth culture, was lower for *S. pneumoniae* than for *E. coli*. This supports a possible role of pneumococcal autolysis in lowering the

ratio of concentration in CFU/mL to concentration in genome-copies/mL, because the autolysed nonviable cells in suspension could contribute to the quantity measured by DNA extraction but not to the quantity measured by colony-counting. Additionally, more than one genome may be present per cell, depending on the phase in the cell division cycle. It is not realistically possible for a single genome-copy to give rise to more than one CFU but ratios of >1 in CFU/genome-copies serve to highlight the difficulty in accuracy with either of the measurement methods. DNA concentration may be underestimated due to inefficiencies in nucleic acid extraction, nucleic acid degradation or sample inhomogeneity or overestimated due to impurities in the extraction process. Colony counting could overestimate or underestimate concentration due to sample inhomogeneity, delays in plating out suspensions or inaccuracies in counting individual colonies.

The lack of a difference between suspension methods for *S. pneumoniae* implies that log phase growth does not sufficiently overcome the problem of autolysed cells in suspension, or that pneumococcal autolysis is not the only problem.

The DNA concentration method of quantification was less variable than the colony counting method. Poor correlation between methods may be due to inherent differences in measuring bacterial cells versus measuring their genomes, rather than simply pneumococcal autolysis as it was more pronounced for *E. coli* than for *S. pneumoniae*. Growing the suspension to log-phase before quantification did not alter the results, but the median ratio of concentrations was lower for *S. pneumoniae* than for *E. coli*, also suggesting that pneumococcal autolysis is not the sole cause of poor correlation between quantification methods.

It was noted that while the DNA concentration method for making standards may appear to be a good choice for quantitative real-time PCR in a single laboratory, that there may be considerable intra-assay variation between batches of standards by either method. What is more, the lack of correlation between methods means that absolute quantification of samples between laboratories using different methods would not give comparable results.

It is important to be aware that the method used for assigning values to quantitative real-time PCR standards will affect the results obtained. This is particularly pertinent when applying thresholds or cut-offs that rely on quantification from an assay designed in one laboratory and then used in another laboratory.

3.3.5 Conclusions

Neither method of preparing in-house standards was considered suitable for the PhD objectives, nor for the Pneumonia Etiology Research for Child Health (PERCH) multi-site study, because they would not be reproducible across the length of each project, nor sufficiently reproducible between laboratories. Instead, it was decided to obtain commercially available plasmids containing the *lytA* target amplicon and serially dilute the plasmids for each individual run (rather than storing diluted plasmids). These plasmids came from Fast-Track Diagnostics (FTD) in Luxembourg. FTD also measure the concentration of the plasmids by Nanodrop, but these are much more pure than whole genomic DNA, thus the measurement is more reliable. Each batch of plasmids that FTD produce is checked to ensure that the cycle thresholds are reproducible within +/- 2 Ct values.

3.4 Matrix for dilution of standards for *lytA* qPCR

The intent was to serially dilute the FTD-supplied *lytA* standards for each *lytA* qPCR run in molecular grade water. Samples of whole blood, of nasopharyngeal swab specimens in STGG or nasopharyngeal swab specimens in viral transport medium that were to be run in the *lytA* qPCR assay would be DNA suspended in eluate from the original sample type, which could have different properties affecting PCR efficiency compared to the standards. In order to ensure this was not the case, FTD *lytA* plasmid standards were serially diluted in molecular grade water, in eluate from whole blood, in eluate from uninoculated STGG and in eluate from uninoculated Copan universal transport medium (UTM). The aim was to compare Ct values of *lytA* qPCR from plasmids serially diluted in sample matrix to plasmids serially diluted in molecular grade water, for each sample type.

3.4.1 Methods

Uninoculated sterile whole blood, STGG and Copan UTM underwent DNA extraction using QIAamp DNA blood mini kits (Qiagen, Germany) as described in the chapter seven for blood or nasopharyngeal swabs. Master stock of *lytA* plasmid standards at 10^8 copies/mL was serially diluted 1:10 four times in molecular grade water and in whole blood eluate and each dilution series were run in triplicate in the same PCR plate. *lytA* qPCR was performed as in section 7.9 of chapter seven and mean Ct values plotted. Master stock of *lytA* plasmid standards at 10^8 copies/mL was serially diluted 1:10 five times in molecular grade water and in STGG and Copan UTM eluates and each dilution series were run in triplicate in the same PCR plate. Mean Ct values were again plotted for each plasmid concentration.

3.4.2 Results

The mean Ct values for lytA plasmid standards in molecular grade water and in whole blood eluate matrix can be seen in figure 3.4.

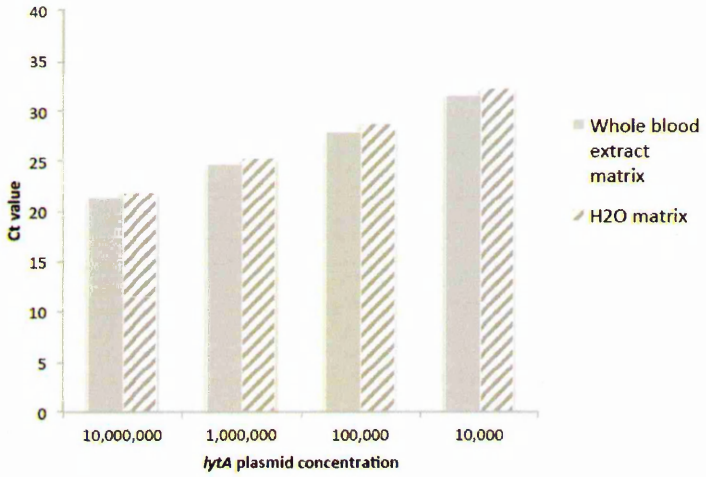


Figure 3.4: Standards matrix. Comparison of molecular grade water and whole blood eluate as matrix for serial dilution of plasmid standards for *lytA* qPCR.

The mean Ct values for *lytA* plasmid standards in molecular grade water and in STGG eluate matrix and Copan eluate matrix can be seen in figure 3.5.

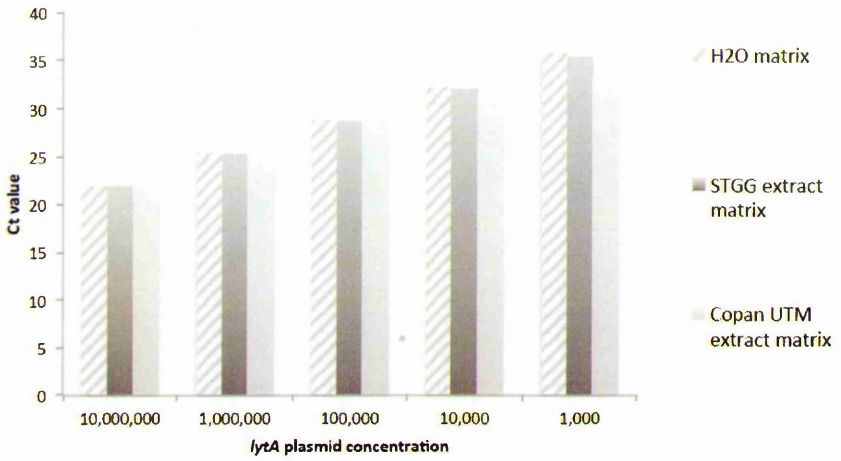


Figure 3.5: Standards matrix. Comparison of molecular grade water, eluate of uninoculated STGG and eluate of uninoculated Copan universal transport medium (UTM) as matrix for serial dilution of plasmid standards for *lytA* qPCR.

3.4.3 Conclusions

Molecular grade water was considered to be a reasonable and convenient replacement for sample extract matrix for serial dilution of standards for use in *lytA* qPCR. It was noted that unsurprisingly there is more variation in Ct values at lower concentrations of the plasmid standard.

3.5 Choosing a template volume for *lytA* qPCR

The original CDC method for *lytA* qPCR used 2.5µL of template in the PCR reaction, for historical reasons (Bernard Beall, personal communication to David Murdoch). Other laboratories have used 6µL[144] or 10µL of template. [147] Increasing the template volume in

the reaction may improve the sensitivity of the assay, but increasing it too much could result in PCR inhibition. An experiment was therefore performed to see which of 2.5µL, 5µL or 10µL would perform best with the *lytA* qPCR protocol.

3.5.1 Methods

Using the *lytA* PCR method as described in section 7.11 in chapter seven, and maintaining the total reaction volume at 25µL, the water was adjusted in the mastermix according to the change in template volume. A set of standards was made by serial 1 in 10 dilution in molecular grade water of DNA extracted from a suspension of *S. pneumoniae*, the concentration measured by the colony counting method (further details can be seen in the chapter seven, section 7.2). The series of standards in triplicate was run for each of the three templates, 2.5µL, 5µL or 10µL, in the same reaction plate.

3.5.2 Results

The mean cycle threshold values for standards run with the three alternative template volumes can be seen in table 3.2.

Table 3.2: Mean Ct values for a series of standards run in triplicate, comparing three alternate DNA template volumes in *lytA* qPCR.

Standards CFU/mL	Mean Ct for 2.5µL template	Mean Ct for 5.0µL template	Mean Ct for 10µL template
10e6	23.06	22.37	23.41
10e5	26.80	26.07	27.87
10e4	30.65	29.66	32.32
10e3	34.56	33.67	35.62
10e2	37.34	36.44	39.11
10e1	negative	37.70	negative

3.5.3 Conclusions

As the 5µL template volume gave lower or equivalent Cycle threshold (Ct) values in the PCR reaction than the 2.5µL or 10µL template volumes, this volume was chosen to continue with. Since the PERCH project chose 10µL as a template volume, and methods needed to be standardised across all seven sites, 10µL was used for the whole blood *lytA* qPCR assay, but a 5µL template volume continued to be used for the nasopharyngeal samples as these were likely to have a higher bacterial load.

3.6 Lower limit of detection and lower limit of linearity of *lytA* qPCR

Once satisfied with the standards, experiments were then performed to ascertain the lower limit of detection of the *lytA* qPCR assay, and to establish the linear range.

3.6.1 Methods

Using the *lytA* PCR method as described in section 7.11 in chapter seven, the plasmid standards were run in triplicate five times, with a fresh 1:10 dilution series made for each run from a master stock vial of 10^8 copies/mL. Two different operators were included. The standard curves were plotted overlaid and the mean and standard deviation of the Ct value for each plasmid concentration plotted with and without the least dilution, of 10^2 copies/mL.

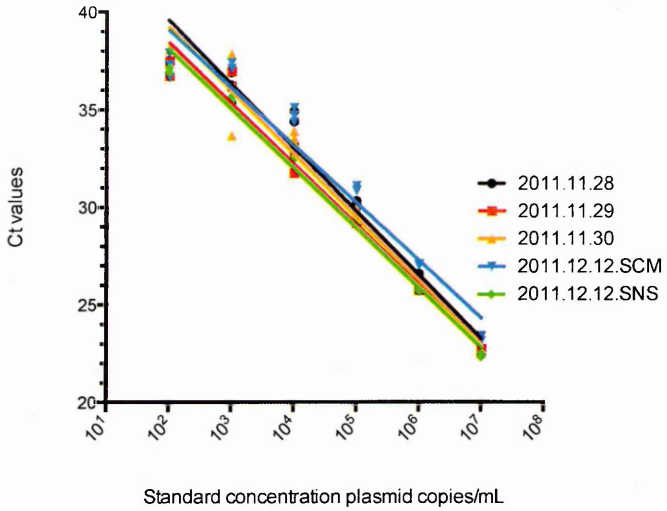
Subsequently, after the project in chapter nine, nasopharyngeal pneumococcal concentration before, during and after respiratory viral infection, was performed, standard curves from another 11 runs of *lytA* qPCR were plotted overlaid and the mean and standard deviation of the Ct value for each plasmid concentration were plotted.

The lower limit of detection was defined as the least *lytA* plasmid concentration that was able to be detected 95% of the time and was extrapolated from these two sets of data. The lower limit of linearity was defined as the least *lytA* plasmid concentration at which the standard curve remained linear.

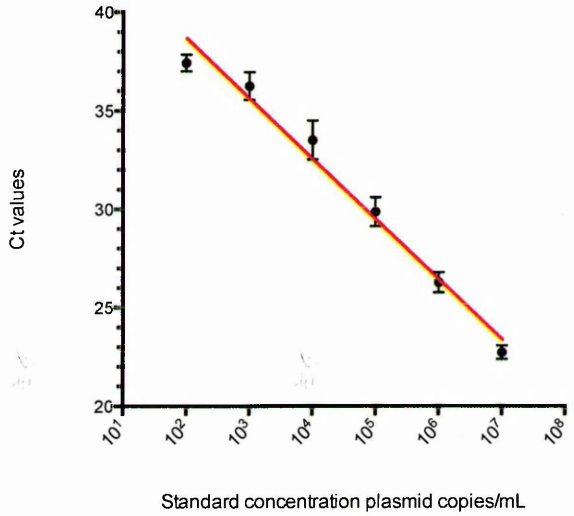
3.6.2 Results and Conclusions

The reproducibility of five individual standard curves with *lytA* plasmids and the mean standard curve with standard deviations about each point can be seen in figure 3.6.

A


Ct values
40
35
30
25
20
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
2011.11.28
2011.11.29
2011.11.30
2011.12.12.SCM
2011.12.12.SNS
Standard concentration plasmid copies/mL

B


Ct values
40
35
30
25
20
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Standard concentration plasmid copies/mL

C

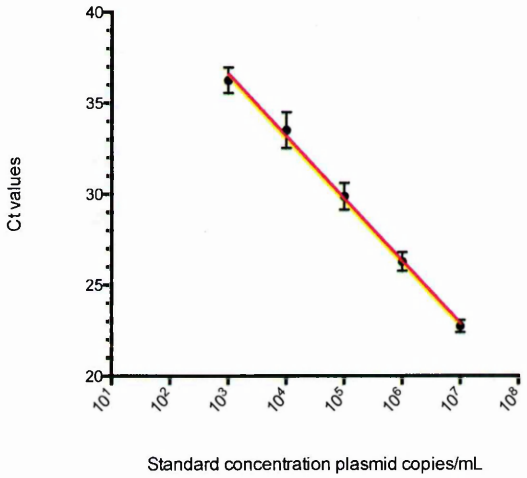

Ct values
40
35
30
25
20
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Standard concentration plasmid copies/mL

Figure 3.6: Reproducibility of *lytA* standard curves. A: Five standard curves using commercially available *lytA* plasmid standards. **B:** Mean standard curve with standard deviations around each point. Slope = -3.1 **C:** Mean standard curve with standard deviations around each point, omitting the 10^2copies/mL standards to demonstrate improved linearity, slope = -3.4.

The slope of the mean standard curve was -3.1 with the 10^2 copies/mL standards included, or -3.4 without these data points. Observation of the standard curves with and without the 10^2 copies/mL data points revealed the lower limit of linearity to be 10^3 copies/mL.

The lower limit of detection should be defined as the concentration at which 95% of the samples can be detected. From these five initial method validation runs, 100% of 15 plasmid standards at 10^4 copies/mL were detected, 80% of 15 plasmid standards at 10^3 copies/mL were detected and 60% of 15 plasmid standards at 10^2 copies/mL were detected. We could therefore take the lower limit of detection to be approximately 10^3 copies/mL or we could extrapolate from these points on a semilog line, which yields a lower limit of detection of 5.62×10^3 copies/mL.

Taking the 11 runs of *lytA* qPCR data from chapter nine, where nasopharyngeal pneumococcal concentration was assessed before, during and after respiratory virus infections, the 33 plasmid standards at 10^4 copies/mL were detected 100% of the time, the 33 plasmid standards at 10^3 copies/mL were detected 85% of the time, and the 33 plasmid standards at 10^2 copies/mL were detected 15% of the time. These standard curves can be seen in figure 3.7. The lower limit of detection of the assay in this dataset

was again approximately 10^3 copies/mL, or by extrapolation $4.64x10^3$ copies/mL. The lower limit of linearity was also 10^3 copies/mL.

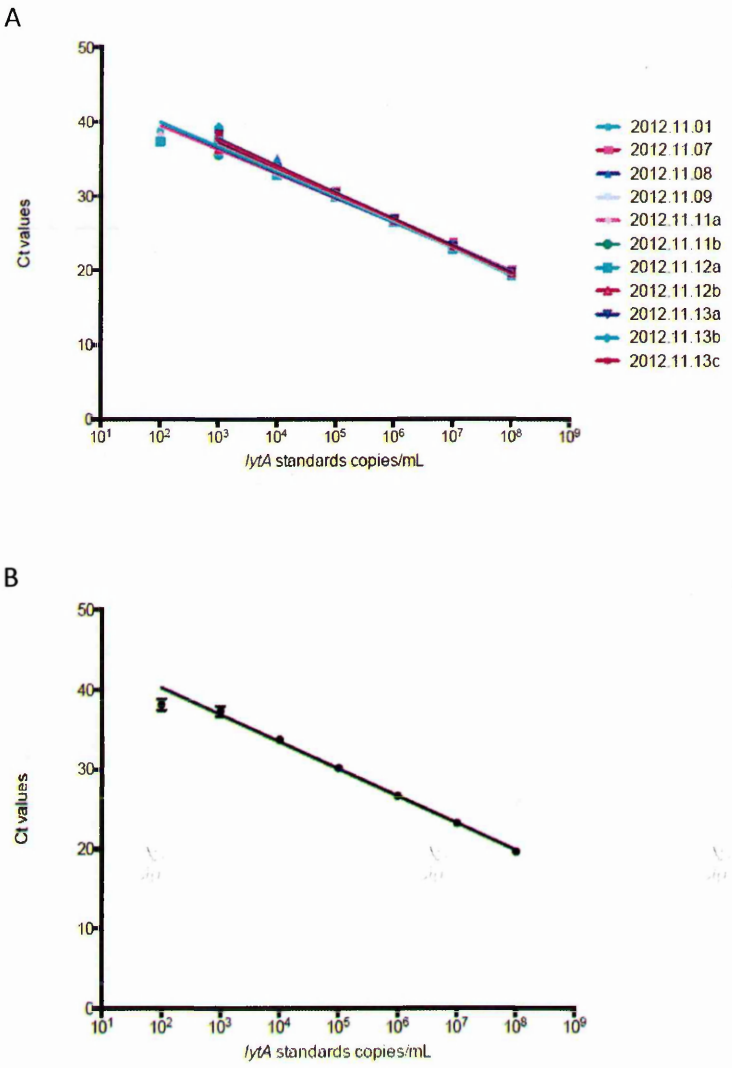


Figure 3.7: Reproducibility of *lytA* standard curves using plasmid standards during a study of pneumococcal concentration from nasopharyngeal swabs in viral transport medium. A: Eleven standard curves using *lytA* plasmid standards. **B:** Mean standard curve with standard deviations around each point.

Chapter Four

4 Validation of laboratory methods for whole blood specimens

4.1 Introduction

Having performed initial method validation for the *lytA* qPCR assay, the next step was to consider the different specimen types that would be used in the upcoming projects. The first specimen type considered was whole blood in EDTA.

Sources of variation in the assay considered included the volume of samples, PCR inhibition, and the DNA extraction method.

4.2 Variation in sample volume

All specimens have had DNA extracted from 200µL of blood, unless there is less than 200µL available, which is sometimes the case with sick infants. In this case the volume extracted was noted and taken into account in the calculation of pneumococcal concentration.

4.3 PCR inhibitors

PCR inhibitors could have been present in some samples. Internal controls could have been spiked into all samples, or a human gene target used, and amplified in multiplex but this could cause amplification competition and reduce sensitivity of the *lytA* qPCR, so after review of the published literature, [116,122,124,125,135,144,145,184] as seen in table 4.1, and discussion with Nicole Wolter from the NICD and David Murdoch it was elected not to do this.

Table 4.1: Use of an internal control in the examination of blood samples for presence of the *lytA* gene

Reference	Country	Specimen type	N	Internal control	Comment
Abdeldaim *et al* CMI 2009 (122)	Sweden	Plasma	183	No	2 samples negative for *lytA* PCR grew pneumococcus in blood culture; samples were spiked and re-tested to check for inhibition, no inhibition detected
Azzari *et al* JMM 2008 (144)	Italy	Whole blood	92	No	
Carvalho *et al* JCM 2007 (125)	USA	Serum	30	Yes	RnaseP
Kee *et al* DMID 2008 (135)	Spain	Whole blood	28	No	Internal amplification control was deliberately not used in order to maximise sensitivity
Rello *et al* Chest 2009 (145)	Spain	Whole blood	93	No	Same group as for Kee *et al* above
Sheppard *et al* JMM 2004 (116)	UK	Whole blood	198	Yes	Bacteriophage lambda sequence amplifiable by lytA primers. 5 samples found to be inhibitory.
Smith *et al* JCM 2009 (184)	UK	Whole blood	332	Yes	Same group as for Sheppard *et al* above. No samples found to be inhibitory.

4.4 The DNA extraction method

The DNA extraction method required validation. It was important early on to establish a DNA extraction technique for whole blood samples because it was thought that sensitivity of pneumococccal PCR from blood would depend significantly on specimen quality and handling, and the DNA extraction method. DNA should be extracted from blood while it is as fresh as possible, prior to freezing, and the extraction method should be good enough to extract even low concentrations of pneumococci.

The options for DNA extraction were constrained by the requirement to share the specimens with other research projects such as the PERCH multi-site international study and the Kilifi genome-wide association studies, and by limited sample volume.

Method validation for DNA extraction first compared the QIAamp DNA blood mini kit (Qiagen, Germany) manual method, with the ABI PRISM 6100 (Applied Biosystems, Life Technologies, California, USA) semiautomated vacuum method being used by the Kilifi genome-wide association project at the time, and the MagNA Pure DNA III (Roche, Indianapolis, USA) automated silica bead-based method (which was being considered by the PERCH multi-site study team, was the method in use for whole blood DNA extraction for *lytA* qPCR at the NICD, and was available in the virology department in the KWTRP laboratory). This work was performed in conjunction with the Kilifi genome-wide association project research team. The work led on to a comparison of the Qiagen manual method with MagNA Pure extraction from pneumococcal spiked whole blood samples, measured by *lytA* qPCR.

4.4.1 Comparison of Qiagen manual method with ABI-PRISM-6100 semi-automated method and MagNA Pure automated method

4.4.1.1 Methods

Details of extraction protocols for Qiagen and MagNA Pure methods can be seen in sections 7.6.1 and 7.4 in chapter seven. The ABI PRISM-6100 method was as follows: 20µL of Proteinase K Solution, 80µL of PK Digestion Buffer and 200µL of whole blood sample was placed in each well of a 96 deep well plate and mixed by pipetting up and down five times. The plate was incubated at 65°C for one hour in a water bath before adding 300µL of BloodPrep DNA Purification Solution for a total volume of 600µL per sample. After mixing by pipetting again, the plate was loaded onto the instrument for a purification run with the settings found in table 4.1.

Table 4.2: Settings for ABI PRISM 6100 semi-automated extraction instrument for DNA extraction from whole blood samples.

Step	Description	Volume (mL)	Position	Incubation (s)	Vacuum (%)	Time (s)
1	Load Samples	600	Waste	0	80	300
2	Add BloodPrep DNA Purification Solution	500	Waste	0	80	400
3	Add BloodPrep DNA Wash Solution	500	Waste	0	80	60
4	Add BloodPrep DNA Wash Solution	500	Waste	0	80	60
5	Add BloodPrep DNA Wash Solution	300	Waste	0	80	60
6	Pre-Elution Vacuum	-	Waste	0	100	60
7	Touch Off	-	Waste	-	-	-
8	Elution Solution 1	80	Collection	600	60	120
9	Elution Solution 2	80	Collection	600	60	120
10	Touch Off	-	Collection	-	-	-

Numbers of whole blood samples in EDTA were as follows, because there was insufficient specimen volume to test every sample by all three methods: ABI n=79, MagNA Pure n=31, Qiagen n=12. Total DNA extracts from all methods were measured by the Nanodrop

spectrophotometer (Thermo Scientific, Delaware, USA) for DNA concentration (in ηg/μL) and purity.

4.4.1.2 Results

The mean and standard deviation of DNA concentrations obtained by the three methods can be seen in table 4.2.

Table 4.3: DNA concentration of extracted whole blood samples by ABI-6100 semi-automated vacuum extractor, Qiagen manual method and MagNA Pure automated extraction instrument.

Extraction method	N	Mean DNA concentration ng/μL	Standard deviation	Range	
ABI-6100	79	51.44	41.21	3.66	150.80
Qiagen mini kit	12	198.86	148.13	48.52	581.00
MagNA Pure	31	14.63	5.62	5.72	31.28

It was noted that some samples clogged in the ABI vacuum system, resulting in high concentration, low volume DNA extracts.

When measuring DNA concentration with the Nanodrop spectrophotometer, the primary indicator of DNA purity is the 260:280 ratio, which should be >1.8. A secondary measure of DNA purity is the 260:230 ratio. The 260:280 and 260:230 ratios of extracted whole blood samples using the ABI-6100 semi-automated vacuum extractor, the Qiagen manual method and the MagNA Pure automated extraction instrument can be seen in table 4.3.

Table 4.4: DNA purity of extracted whole blood samples by ABI-6100 semi-automated vacuum extractor, Qiagen manual method and MagNA Pure automated extraction instrument, as measured by the Nanodrop 260:280 and 260:230 ratios.

Extraction Method	N	Mean 260/280 ratio	Standard deviation	Range	
ABI-6100	79	1.9	0.71	1.41	5.60
Qiagen mini kit	12	1.85	0.06	1.68	1.91
MagNA Pure	31	2	0.19	1.66	2.56
Extraction Method	**N**	**Mean 260/230 ratio**	**Standard deviation**	**Range**	
ABI-6100	79	0.19	0.13	0.02	0.57
Qiagen mini kit	12	1.43	0.36	0.98	2.15
MagNA Pure	31	0.64	0.16	0.35	1.09

4.4.1.3 Conclusions

The Qiagen manual method gave the greatest DNA concentration as measured by Nanodrop, but is not a high-throughput method. The MagNA Pure method gave the most pure DNA as measured by Nanodrop. The ABI method gave a good DNA concentration most of the time but tended to clog meaning some samples were lost. The Qiagen method appeared to be best so far, and it was decided to compare the *lytA* qPCR results from DNA extracted by Qiagen compared to MagNA Pure as a next step.

4.4.2 Comparison of Qiagen manual method with MagNA Pure automated method using pneumococcal spiked whole blood samples, measured by *lytA* qPCR

4.4.2.1 Methods

A heavy (McFarland 3.0) suspension of *Streptococcus pneumoniae* ATCC 49619, was spiked into whole EDTA blood and serial 1:10 dilutions were performed, each dilution was extracted by both MagNA Pure and Qiagen methods before undergoing *lytA* PCR in

duplicate, and had colony counting performed. Chapter seven, sections 7.4, 7.6.1 and 7.11, has details for the MagNA Pure and Qiagen extraction protocols and the *lytA* PCR method. The *lytA* PCR method in use was as described for bacterial isolates.

4.4.2.2 Results

The Ct values of *lytA* PCR performed on duplicate spiked whole blood specimens extracted by the Qiagen manual method and the MagNA Pure automated instrument can be seen in table 4.4.

Table 4.5: Ct values from *lytA* real-time PCR of whole blood extracted by the Qiagen manual method or the MagNA Pure automated method.

Colony counting CFU/mL	Ct from Qiagen extract	Duplicate from Qiagen	Ct from MagNA Pure extract	Duplicate from MagNA Pure
10e0	negative	negative	negative	negative
10e1	negative	negative	negative	negative
10e2	37.45	negative	negative	negative
10e3	35.75	34.65	negative	negative
10e4	31.27	31.21	33.14	32.41
10e5	27.39	27.66	29.16	29.16
10e6	23.89	23.89	24.69	24.69
10e7	21.54	21.56	21.57	20.98

4.4.2.3 Conclusions

The limit of detection using the *lytA* PCR was 2 log higher from MagNA Pure-extracted DNA than from Qiagen-extracted DNA, so the Qiagen method appeared to be more sensitive in combination with the *lytA* PCR than the MagNA Pure method.

At that point in time whole blood specimens being collected on all children who were admitted to Kilifi County Hospital and qualified for a blood culture, began being extracted using the QIAamp DNA blood mini kit manual method, within three days of collection of the specimen.

We next compared the manual Qiagen extraction to automated QIAxtractor extraction – also from Qiagen, and using the same chemistry as the manual Qiagen method. The experiments included the QIAxtractor as being considered by the genome-wide association studies research group for high throughput extraction of frozen stored blood samples and the EasyMag extraction method because a subset of these samples were included in the PERCH multi-site study which by that stage required automated EasyMag extraction as a standard method across sites.

4.4.3 Comparison of manual Qiagen method with QIAxtractor and EasyMag automated methods, using pneumococcal spiked whole blood samples.

The aim was to compare the three extraction methods' ability to extract pneumococcus from whole blood at a range of concentrations. A secondary objective was to compare robustness of the extraction methods using fresh and one month-old blood samples.

4.4.3.1 **Methods**

EDTA whole blood was spiked with a 0.5 McFarland suspension of *Streptococcus pneumoniae* ATCC 49619. This was serially diluted 1:10 in fresh whole blood, and extracted in duplicate by each of the EasyMag (Bio-Merieux, France) and QiaXtractor (Qiagen, Germany) automated instruments and by the Qiagen manual method using the QIAamp blood minikit (Qiagen, Germany) (Details of the extraction protocols can be seen in chapter seven, sections 7.7, 7.5 and 7.6.1). Extracted DNA was stored at -80°C. Aliquots of each spiked sample were retained for one month at 2-8°C and were subsequently extracted again by each method, also in duplicate, to check for robustness of the method with older samples. *lytA* PCR was performed according to the description in section 7.9 in chapter seven.

4.4.3.2 **Results**

The *lytA* PCR Ct values of aliquots of pneumococcus-spiked blood samples from which DNA was extracted by three different methods while samples were fresh and again after one-month of storage can be seen in figure 4.1.

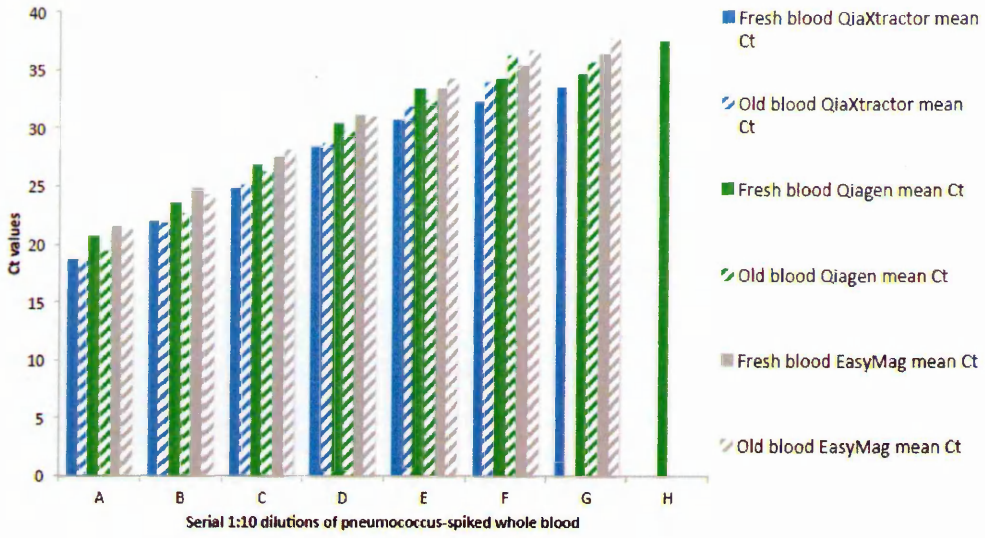


Figure 4.1: *lytA* PCR Ct values of pneumococcus-spiked blood samples processed fresh and after one month at 2-8°C

4.4.3.3 **Conclusions**

Ct values of *lytA* qPCR from pneumococcal-spiked whole blood samples were comparable by the three extraction methods; manual Qiagen, automated QIAxtractor and EasyMag. QIAxtractor gave the lowest Ct values, followed by the manual Qiagen method, followed by EasyMag. All methods returned linear results over the range of dilutions. DNA extracted from fresh samples performed better in the *lytA* PCR than DNA extracted from samples stored at 2-8°C for one month only in the lowest concentrations.

It was notable that the QIAxtractor method uses a 96 well plate and consumables which are wasted if less than 96 samples are run at a time, which is a problem for extracting from fresh, rather than batched frozen, samples.

The fresh whole blood samples from all admissions therefore continued to be extracted by the manual Qiagen method, within 3 days of sample collection. Blood samples from patients also enrolled in the PERCH study additionally underwent EasyMag extraction, but this was from frozen blood samples most of the time as the EasyMag instrument had technical problems and had to be replaced.

Chapter Five

5 Validation of laboratory methods for nasopharyngeal swabs in viral transport medium

5.1 Introduction

Potential sources of variation in *lytA* concentration in nasopharyngeal swab samples are high compared to blood specimens. DNA extraction, possible PCR inhibition and the volume of posterior nasopharyngeal secretions were considered as sources of variation. Reproducibility of results was assessed.

5.2 DNA extraction method

For pragmatic reasons, the total nucleic acid (TNA) extracts of the viral transport medium (VTM) specimens that had been performed in the parent study of respiratory virus transmission[9] were used for objective two, the study of pneumococcal concentration before, during and after respiratory virus infection. The method that had been used in the parent study was the MagNA Pure (Roche, Indianapolis, USA) instrument, an automated magnetic silica bead extraction method, with the MagNA Pure High Performance Total Nucleic Acids kit, [8] to extract nucleic acids from Copan nasopharyngeal flocked swabs in an in-house VTM. The VTM had been made from 5g veal infusion broth (BD Difco, USA), 1g bovine albumin fraction V (MP Biomedicals, France), 20mg gentamicin, 400µg amphotericin in 200ml of sterile distilled water. Reproducibility of pneumococcal DNA amplification from these samples was assessed for the present project using unneeded participant samples from the parent study.

For the study of the performance of *lytA* qPCR on whole blood, nasopharyngeal pneumococcal concentration was assessed as part of the Pneumonia Etiology Research for Child Health (PERCH) project. These nasopharyngeal flocked swabs were collected into Copan universal transport medium (UTM) (Copan Diagnostics, California, USA) and extracted using the EasyMag automated extraction instrument (Bio-Merieux, France).

The aim was to ensure both EasyMag and MagNA Pure LC instruments could successfully extract known quantities of pneumococcus from the transport medium used for nasopharyngeal swabs that had been collected for viral detection.

5.2.1 Methods

Streptococcus pneumoniae ATCC 49619 was grown on blood agar overnight and a suspension made in phosphate buffered saline at an optical density of 0.12 (Biowave colorimeter, Biochrom, UK), estimating a concentration of $10^7 - 10^8$ cells/mL. 140µL of the suspension was added to 1260µL of Copan UTM to make a 1:10 dilution, and a further five 1:10 dilutions made in UTM, resulting in six different concentrations labelled A to F. These were aliquoted such that duplicates of each concentration underwent EasyMag extraction and duplicates of each concentration underwent MagNA Pure extraction. EasyMag extractions were performed from 400µL of spiked UTM into 110µL of eluate, and MagNA Pure extractions were performed from 200µL of spiked UTM into 60µL of eluate, to mimic the PERCH and household viral transmission study methods from which it was expected study samples would be drawn. All nucleic acid extracts underwent *lytA* qPCR, as described in section 7.10 of chapter seven.

5.2.2 Results

Both EasyMag and MagNA Pure were able to extract pneumococcus from UTM to an estimated concentration of between 10^3 to 10^4 cells/mL. The mean *lytA* Ct values from the duplicate reactions, and the mean quantity of *lytA* detected adjusted for extraction:elution ratio, for each extraction method, can be seen in table 5.1 and figure 5.1.

Table 5.1: *lytA* qPCR results from EasyMag and MagNA Pure extractions of pneumococcus-spiked Copan universal transport medium samples

Spiked serially diluted sample	Expected quantity cells/mL	EasyMag mean Ct	EasyMag mean quantity	MagnaPure mean Ct	MagnaPure mean quantity
A	10,000,000	23.8	2,041,853	23.7	2,312,376
B	1,000,000	26.8	362,808	27.4	262,896
C	100,000	29.5	65,438	30.8	33,460
D	10,000	34.1	4,280	34.6	4,121
E	1,000	36.6	968	36.9	878
F	100			37.4	651

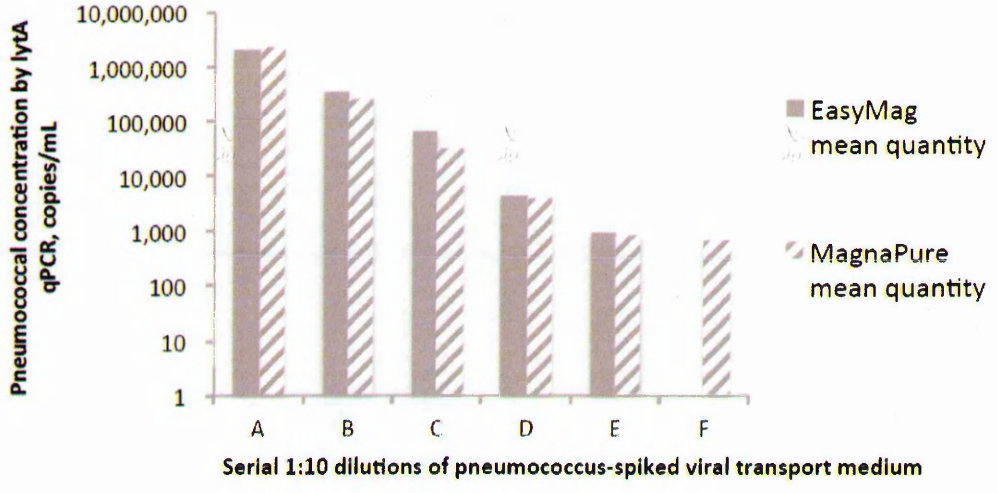


Figure 5.1: Pneumococcal concentration in spiked Copan transport medium, by *lytA* qPCR after EasyMag and MagNA Pure DNA extraction

5.2.3 Conclusions

EasyMag and MagNA Pure extraction methods for pneumococcal DNA from UTM were equivalent within the range of $10^7 - 10^3$ cells/mL. Concentrations obtained by either method were slightly less than the approximate expected concentrations, but colony counting had not been performed to confirm the expected concentrations.

5.3 The possibility of PCR inhibitors present in nasopharyngeal swab samples

It is possible that nasopharyngeal swabs in transport medium contain PCR inhibitors inherent to the sample type itself, which cannot be sufficiently tested by spiked samples. This inhibition can usually be overcome by diluting the sample, which dilutes the inhibitors. Therefore actual samples from the parent viral transmission study used in objective two were tested to check for the presence of sample inhibition. If sample inhibition is present then the diluted sample should have the same or lower Ct value than the undiluted sample, and the amplification curve will be more efficient. It was not possible to do this for all samples, due to limitations of sample volume, so a subset was tested.

5.3.1 Methods

Seventeen nucleic acid extracts from the household viral transmission study that were to be used in the pilot study for objective three were selected as a convenience sample. These samples underwent *lytA* qPCR neat and diluted 1:10 in molecular grade water, as described in section 7.10 in chapter seven.

5.3.2 Results

Fifteen of 17 samples were positive for *lytA*. On dilution, Ct values increased for 14 of 15 samples that were positive for *lytA*, and one sample became negative. These can be seen in figure 5.2.

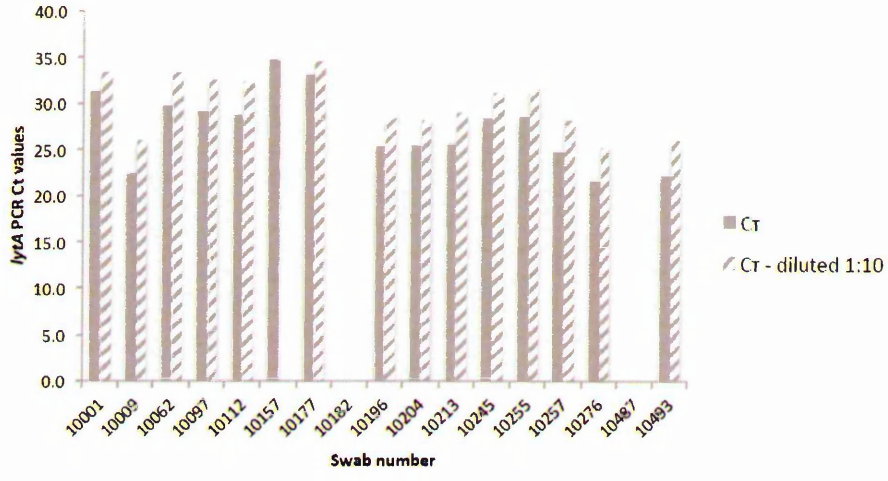


Figure 5.2: *lytA* qPCR Ct values of nasopharyngeal swab samples in viral transport medium tested neat and diluted 1:10

5.3.3 Conclusions

No evidence of *lytA* qPCR sample inhibition was noted; all Ct values increased except for one that became negative with dilution. Dilution of all nasopharyngeal swab samples from the parent viral transmission study before running the *lytA* qPCR was not recommended because some samples that were actually positive might have become negative on dilution.

5.4 The volume of posterior nasopharyngeal secretions sampled

In attempting to measure the nasopharyngeal pneumococcal concentration using a nasopharyngeal swab, we are only estimating the true value. In reality all we can measure is the pneumococcal concentration in a vial of transport medium. During processing of a nasopharyngeal swab collected into transport medium, the sample is vortexed and then nucleic acids are extracted from a standard volume of transport medium and eluted into a standard volume of eluate. The quantitative PCR then measures the concentration of pneumococcus present in the eluate. This concentration can be multiplied by the ratio of the eluate to extraction volume, to give the concentration of pneumococcus in the transport medium, in copies/mL. This concentration would then need to be multiplied by the dilution factor of nasopharyngeal secretion in the transport medium to get the concentration of pneumococcus in the nasopharyngeal secretion; this value is not known.

The nasopharyngeal swab can be collected according to a standard method [168] and some groups have taken a pragmatic approach of assuming a standard volume of nasopharyngeal secretion is collected into the transport medium, for example 100µL of nasopharyngeal secretion from a nasopharyngeal flocked swab.[185] In reality there will be individual variation in the quality and volume of the specimens, both due to the collector's technique and to the study participant's degree of coryza. If the swab is not collected from the actual posterior nasopharynx, but from the anterior or mid-nares, the specimen will not be of adequate quality. If the swab is collected too quickly, it may not contain an adequate volume. Any of these problems would lead to less posterior nasopharyngeal secretion volume in the transport medium. If a study participant has copious coryza, then it is possible the volume put into the transport medium could be greater than expected.

To account for this problem in measuring the pneumococcal concentration in nasopharyneal secretions, the measurement could be adjusted using a marker of sampling variation. One way to achieve this is to measure *lytA* DNA relative to the concentration of DNA from human cells. Quantitative PCR for a gene present in human nasopharyngeal squamous epithelial cells would provide an estimate of the concentration of human cells present in the sample. Since the nasopharyngeal flocked swab is designed to collect posterior nasopharyngeal epithelial cells in order to detect cell-associated virus, this would be a good measure of the adequacy of sample collection. Potential limitations of this method include amplification competition in a multiplex quantitative PCR, so that the target of one set of primers is preferentially amplified. Performing the qPCRs in singleplex would avoid this.

A study carried out by swab manufacturers Copan, on adult volunteers, demonstrated that the quantity of human beta-actin DNA present in a nasopharyngeal flocked swab (NPFS) specimen correlated with the quantity of epithelial cells present, counted directly with immunofluorescent stain.[186] Other groups have used other reference genes such as albumin [187] or RNAseP,[188] instead of beta-actin,[189] to detect presence of human DNA in a clinical sample. A multi-copy human gene, *alu*,[190] was recommended on a European Molecular Biology Laboratory (EMBL) quantitative PCR course (http://www.embl.de/training/events/2012/MIQ12-01/) for this purpose, because of the sensitivity of the assay (personal communication, Jim Huggett, LGC, London, UK).

Ideally, the method should be compared against a gold standard. The gold standard would be a measure of the true pneumococcal concentration on the posterior nasopharyngeal mucosa, which is difficult to achieve. Instead we hypothesised that quality of nasopharyngeal sampling would affect the human DNA quantity obtained from

the sample, so that human DNA concentration would increase as quality of sampling increases. A secondary hypothesis was that the protein concentration measured spectrophotometrically should correlate with the human DNA quantity by quantitative PCR.

5.4.1 Methods

Six adult volunteers were sampled from the nasopharynx three times in succession using nasopharyngeal flocked swabs in Copan universal transport medium (Copan Diagnostics, California, USA). The first swab sampled the anterior nares only, the second swab sampled the mid-nasopharyngeal cavity on the other side, and the third swab sampled the posterior nasopharynx on the first side of the nose. These swabs were labelled a, b and c respectively. In the laboratory, each sample was aliquoted for extraction firstly without any vortexing, then briefly vortexed and aliquoted again, and finally vortexed for 10 seconds before the final aliquot. Each aliquot was labelled x, y and z respectively. 200µL from each sampling condition was extracted using the MagNA Pure High Performance Total Nucleic Acid kit (Roche, Indianapolis, USA) and eluted into 100µL. These nucleic acid extracts were then tested using the *alu* PCR for human DNA quantification and the DNA quantity compared to the quality of the nasopharyngeal sampling. In addition, each aliquot underwent total protein measurement using the Nanodrop (Thermo Scientific, Delaware, USA) and a bovine serum albumin control (Biolabs, Maryland, USA).

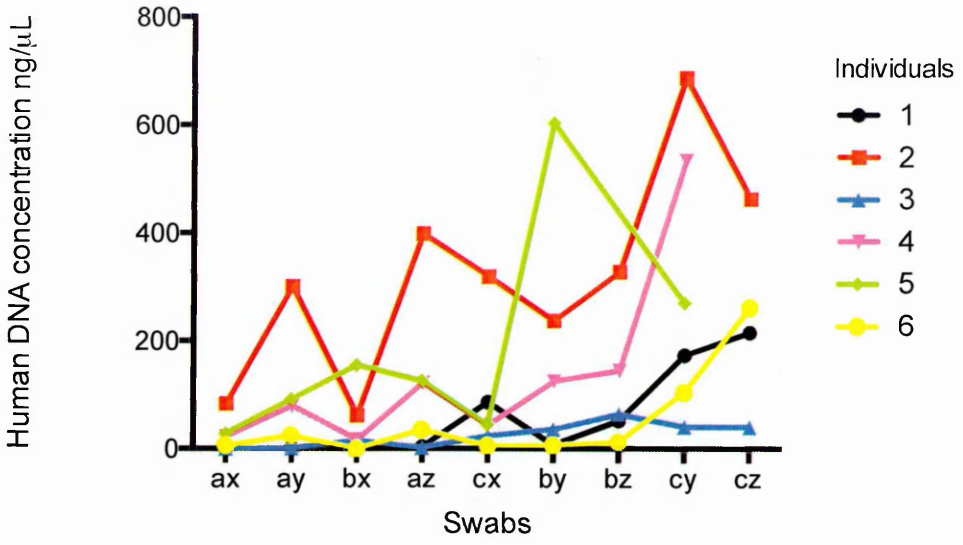
To choose a human DNA qPCR to account for the potential variability in sampling of the posterior nasopharynx a beta-actin real-time qPCR (personal communication, Clayton Onyango, KWTRP) was initially tested and compared with the *alu* real-time qPCR [190] and the *alu* PCR was carried on because it gave more efficient amplification curves.

The *alu* qPCR method used was as described in section 7.12, chapter seven, except that for this experiment the quantity of *alu* detected from each sample in the plate was multiplied by 100/200 (to account for the extraction:elution ratio) and samples were not diluted 1:1000. If samples gave inefficient amplification curves they were repeated after a 1:10 dilution of the nucleic acid extract and if results were similar, omitted from subsequent analysis.

The DNA concentration as measured by *alu* qPCR underwent natural log transformation and the difference between the highest quality swab (deepest swab with longest vortex time) and the lowest quality swab (anterior nasal swab with no vortex) was calculated for each participant. A one-sample t-test was used to compare the mean difference in natural log DNA concentration to zero for the change in swab quality. Spearman's correlation was applied to the human DNA concentration by *alu* qPCR and the total protein concentration of the sample.

5.4.2 Results

The DNA concentration as measured by *alu* qPCR increased as swab quality improved, for all six participants. This can be seen in figure 5.3. The mean change in DNA concentration from the lowest to highest quality nasopharyngeal swab collection was a 35-fold increase (95% confidence interval 7-183 fold increase), p=0.003. Figure 5.4 shows Spearman's correlation co-efficient between DNA concentration as measured by *alu* qPCR and the total protein concentration of the sample.



a = swab in anterior nares only
b = swab in mid nasopharynx
c = swab in posterior nasopharynx

x = no vortexing before DNA extraction
y = brief vortex before DNA extraction
z = prolonged vortex before DNA extraction

Figure 5.3: Human DNA concentration by *alu* qPCR as swab quality increases among adult volunteers

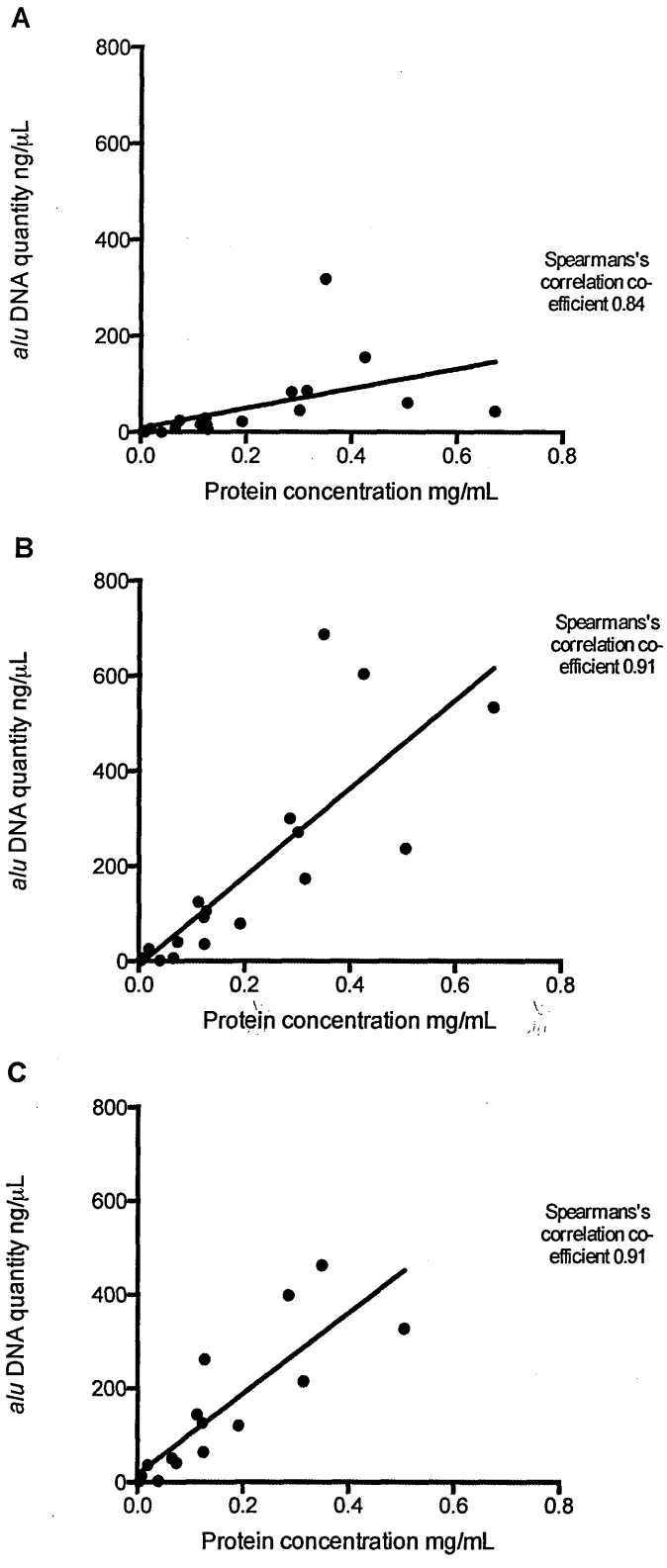

A
800
600
400
200
0
alu DNA quantity ng/μL
0.0
0.2
0.4
0.6
0.8
Protein concentration mg/mL
Spearmans's correlation co-efficient 0.84
B
800
600
400
200
0
alu DNA quantity ng/μL
0.0
0.2
0.4
0.6
0.8
Protein concentration mg/mL
Spearmans's correlation co-efficient 0.91
C
800
600
400
200
0
alu DNA quantity ng/μL
0.0
0.2
0.4
0.6
0.8
Protein concentration mg/mL
Spearmans's correlation co-efficient 0.91

Figure 5.4: Correlation between human DNA concentration by *alu* qPCR and protein concentration measured by Nanodrop spectrophotometer, by time spent vortexing before DNA extraction. A: no vortexing. **B:** brief vortex. **C:** prolonged vortex.

5.4.3 Conclusions

The quantity of DNA in the nasopharyngeal swab sample in transport medium as measured by *alu* qPCR, did increase with the quality of sample collection as hypothesised. The *alu* qPCR measurement correlated better with the total protein concentration in the vortexed samples than the non-vortexed samples. The *alu* qPCR measurement is likely to be a reasonable measure of the nasopharyngeal sample present in the transport medium with which to adjust the pneumococcal concentration to allow for sampling variability. A limitation of this experiment is that none of the adult volunteers had purulent nasal secretions, which might be expected to impact on results of the *alu* qPCR in real conditions when studying children.

Further method validation was planned for *alu* qPCR before use in the main study objectives; to improve PCR efficiency by removing potential PCR inhibitors by diluting the samples before performing *alu* qPCR, and to assess the reproducibility of *alu* qPCR.

5.4.4 Testing sample dilution for *alu* qPCR

The aim was to select the best sample dilution ratio for *alu* qPCR.

5.4.4.1 Methods

5.4.4.1.1 Experiment One: 1:10 vs. 1:100 dilutions

Thirty-two nasopharyngeal swab samples from the respiratory viral transmission parent study for objective two of this thesis that had been extracted by MagNA Pure with the High Performance Total Nucleic Acid kit (Roche, Indianapolis, USA) from 200uL of sample into 60uL of eluate were used. A new buffy coat extract was made (details can be found in section 7.6.4 of chapter seven) and the DNA concentration measured using Nanodrop, to get a higher concentration of total human DNA in the standards, with seven 1:10 serial dilutions from 16.1ηg/μL. The 32 samples were each run at a 1:10 and a 1:100 dilution, by diluting the nucleic acid extracts in molecular grade water. *alu* PCR primers were as previously described in section 7.12, chapter seven, but were used in a mastermix of 0.75μL forward primer, 0.75μL of reverse primer, 6.25μL of SYBR green (Applied Biosytems, Life Technologies, California, USA), 2.25μL of molecular grade water and 2.5μL of template DNA per reaction. Quantitative standards were run in triplicate as were negative template controls and a negative extraction control. Cycling conditions were as described in the section 7.12, chapter seven. Quantitative results were adjusted for dilution factor and for the extraction ratio.

5.4.4.1.2 Experiment Two: 1:100 vs. 1:1000 dilutions

Twenty unneeded nasopharyngeal swab samples in VTM from the parent respiratory viral transmission study used objective two were extracted by MagNA Pure with the High Performance Total Nucleic Acid kit (Roche, Indianapolis, USA) from 200μL of sample into 60μL of eluate. There was insufficient sample volume to use the same samples as in experiment one. The standards and controls were as in experiment one above. The 20

samples were run at a 1:100 and a 1:1000 dilution, by diluting the nucleic acid extracts in molecular grade water. *alu* qPCR was carried out as for experiment one above.

5.4.4.2 **Results**

5.4.4.2.1 Experiment One: 1:10 vs. 1:100 dilutions

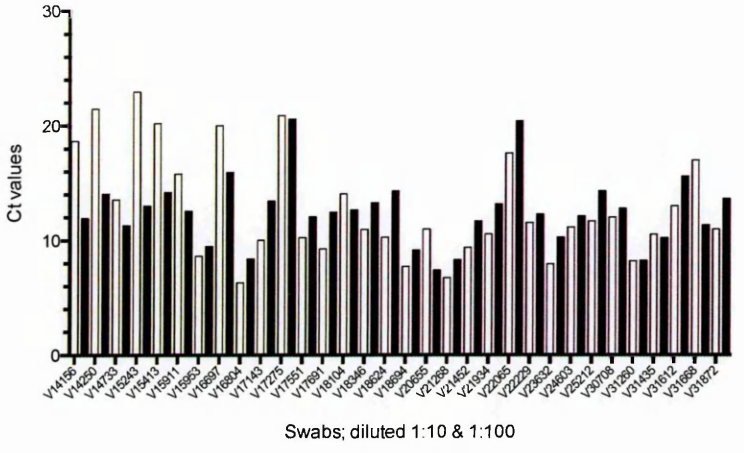
For each pair of results the Ct value was either the same, was shifted to the right as expected for a 1:10 dilution factor (a 1:10 dilution factor should shift a Ct value approximately three cycles to the right), or the Ct value was shifted to the left with an improvement in the efficiency of the amplification curve.

5.4.4.2.2 Experiment Two: 1:100 vs. 1:1000 dilutions

All except one sample (which was unchanged) had a shift in Ct value to the right, with a mean shift of 2.25 cycles.

Figure 5.5 shows the graphs of the Ct values and the DNA concentration for each pair of results comparing 1:10 to 1:100 and 1:100 to 1:1000 dilutions.

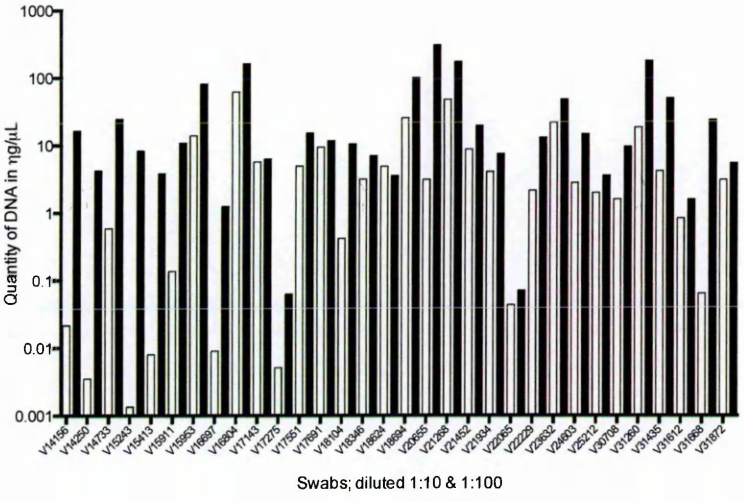
A *alu* qPCR Ct values for swabs diluted 1:10 and 1:100



Swabs; diluted 1:10 & 1:100

White bars = 1:10 dilution
Black bars = 1:100 dilution

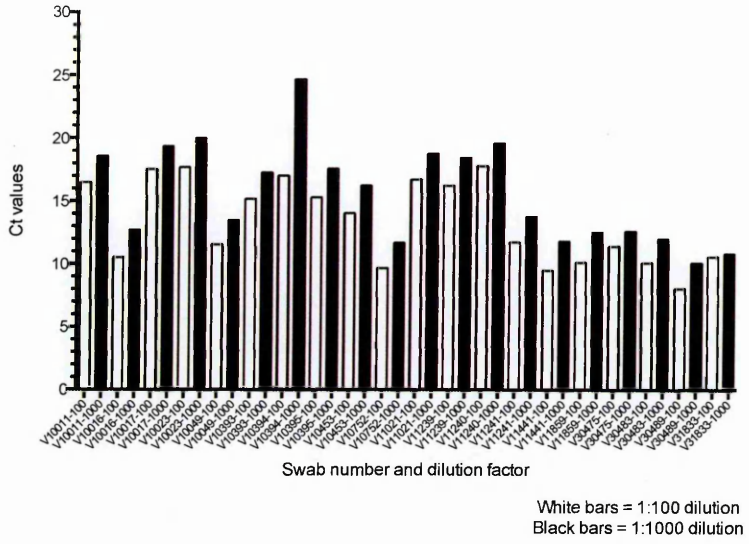
B Human DNA concentration by *alu* qPCR



Swabs; diluted 1:10 & 1:100

White bars = 1:10 dilution
Black bars = 1:100 dilution

C *alu* qPCR Ct values for swabs diluted 1:100 and 1:1000



D Human DNA concentration by *alu* qPCR

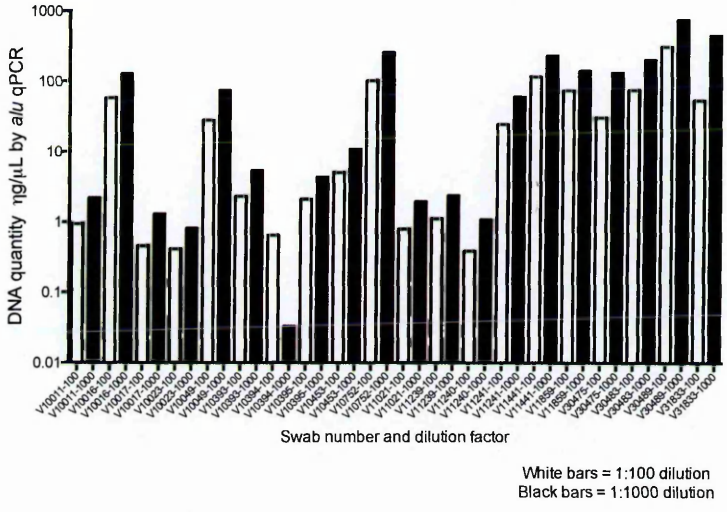


Figure 5.5: Comparing Ct values and human DNA concentration by *alu* qPCR from DNA extracted from nasopharyngeal swab samples after 1:10, 1:100 and 1:1000 dilutions. A: *alu* qPCR Ct values for 1:10 compared to 1:100 dilutions. **B:** Human DNA concentration obtained for 1:10 compared to 1:100 dilutions. **C:** *alu* qPCR Ct values for 1:100 compared to 1:1000 dilutions. **D:** Human DNA concentration obtained for 1:100 compared to 1:1000

dilutions. Swabs in Experiment 2 (C and D) are different than those in Experiment 1 (A and B) due to sample volume limitations.

5.4.4.3 **Conclusions**

For all samples at least a 1:100 dilution is recommended but for some samples a 1:1000 dilution is needed. This improves the efficiency of the PCR by diluting out sample inhibitors and by not overwhelming the reaction with template. Because the *alu* target is found thousands of times in the human genome, this does not result in samples becoming negative and means that the *alu* PCR is likely to be very sensitive to even very small amounts of human specimen being present in sample transport medium.

5.4.5 Reproducibility of *lytA* qPCR and *alu* qPCR

Before commencing work on objective two the reproducibility of *lytA* and *alu* qPCRs on nasopharyngeal swab samples was assessed.

5.4.5.1 **Methods**

The same 20 unneeded samples from the parent study of respiratory virus transmission as in experiment two of section 5.4.4.1.2 above were used, extracted on three different days by three different laboratory technologists, using the MagNA Pure LC instrument with the High Performance Total Nucleic Acid kit (Roche, Indianapolis, USA). This was to ensure the replicates would be true extraction replicates, and not just PCR replicates.

5.4.5.1.1 Experiment Three:

The samples were diluted 1:1000 in a two-step procedure and the *alu* qPCR run as per the method outlined in section 5.4.4.1.1 above. The replicate sets were labelled A, B and C, and each set was run on a separate plate, with separate mastermix, a separate serial dilution of the standard and on a separate ABI-7500 real-time PCR cycler. In addition, the *lytA* qPCR was run on each replicate set, also with separate plates, mastermix, standard dilution and cycler. The *lytA* qPCR method was as outlined in section 7.10 in chapter seven.

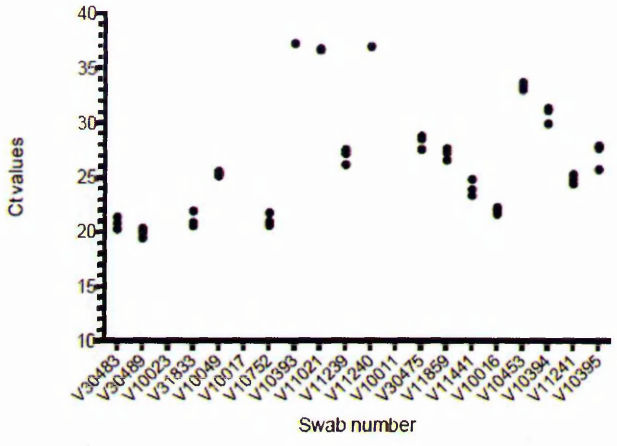
5.4.5.1.2 Experiment Four:

In an effort to reduce variation introduced by the *alu* qPCR step, the *alu* qPCR method was modified slightly from experiment three above, in that the 1:1000 dilutions were performed in a single step and the reaction mix volume was doubled. The *lytA* qPCR was not repeated.

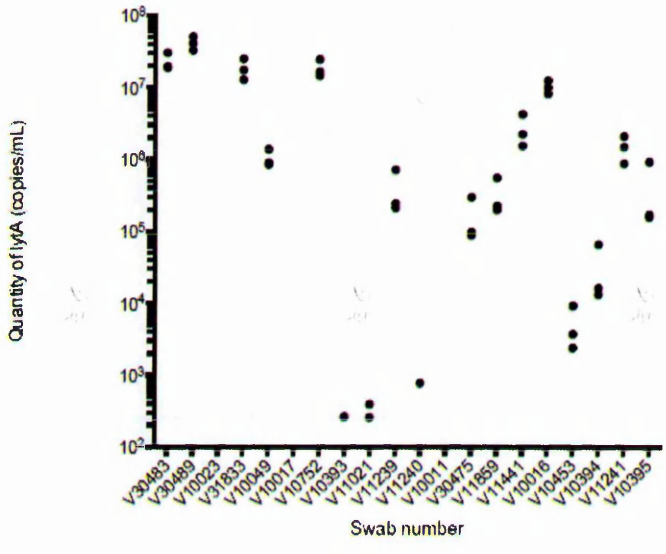
5.4.5.2 **Results**

The *lytA* Ct values and pneumococcal concentration detected in the three replicates, and the mean and standard deviation of the Ct values and of the log transformed concentrations of the nasopharyngeal samples can be seen in figure 5.6.

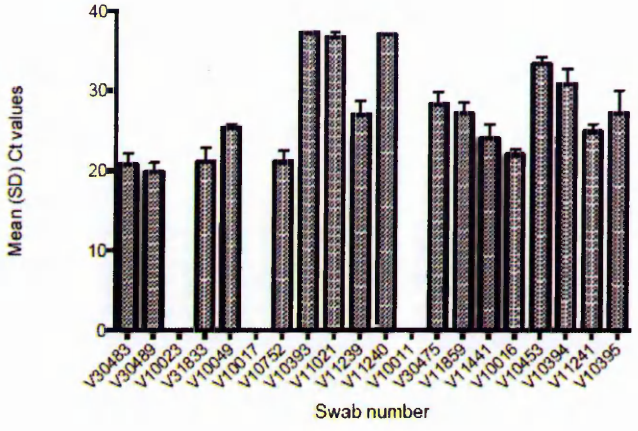
A


Ct values
40
35
30
25
20
15
10
V30483
V30489
V10023
V31833
V10049
V10017
V10752
V10393
V11021
V11239
V11240
V10011
V30475
V11859
V11441
V10016
V10453
V10394
V11241
V10395
Swab number

B


Quantity of lytA (copies/mL)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
V30483
V30489
V10023
V31833
V10049
V10017
V10752
V10393
V11021
V11239
V11240
V10011
V30475
V11859
V11441
V10016
V10453
V10394
V11241
V10395
Swab number

C



D

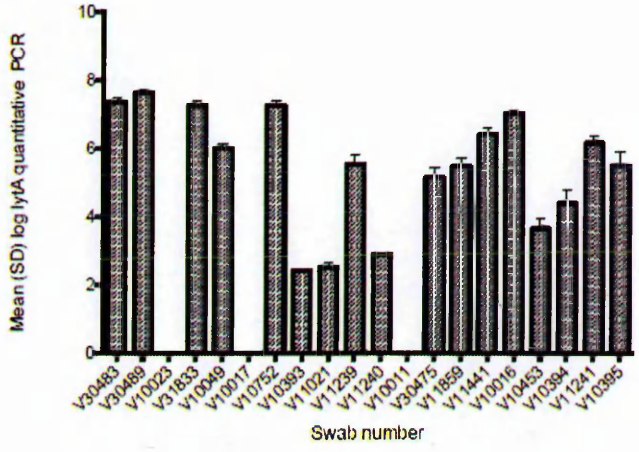
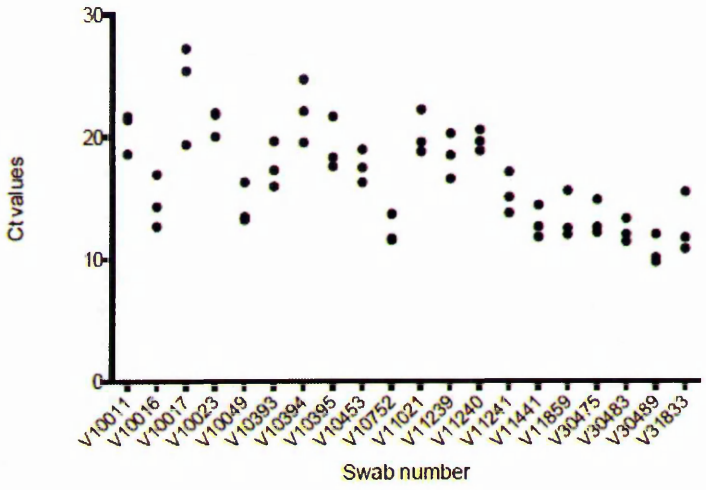


Figure 5.6: Reproducibility of *lytA* qPCR. Three extraction replicates of nasopharyngeal flocked swabs in viral transport medium. Experiment 3. **A**: *lytA* qPCR Ct values. **B:** *lytA* concentration in copies/mL. **C:** Mean and standard deviation of *lytA* Ct values. **D:** Mean and standard deviation of $\log_{10}$ *lytA* concentration in log copies/mL

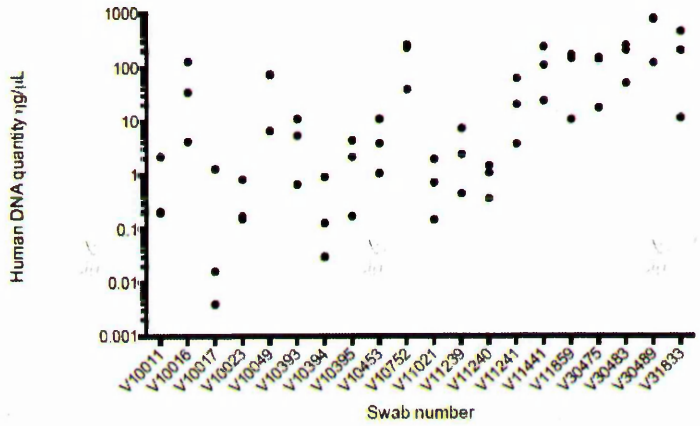
The *alu* Ct values, DNA concentration measured by *alu* qPCR, mean and standard deviation of log transformed human DNA concentration, *lytA* copies/µg human DNA and

mean and standard deviation of log transformed *lytA* copies/µg human DNA, for experiments three and four, can be seen in figure 5.7.

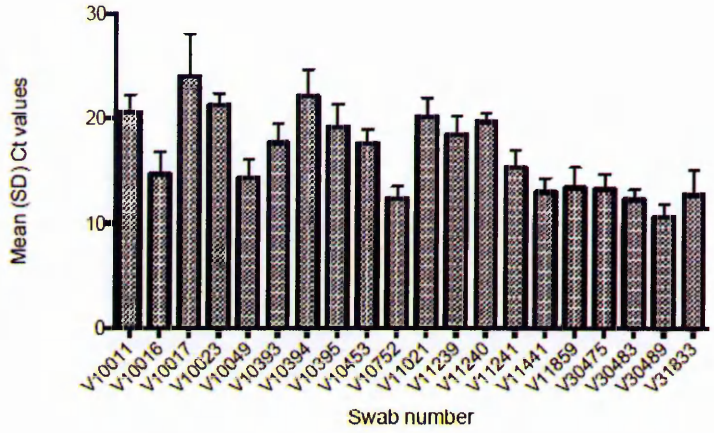
A Reproducibility of *alu* qPCR Ct values



B Human DNA concentration



C Mean and standard deviation of *alu* qPCR Ct values



D Mean and standard deviation of natural log human DNA concentration

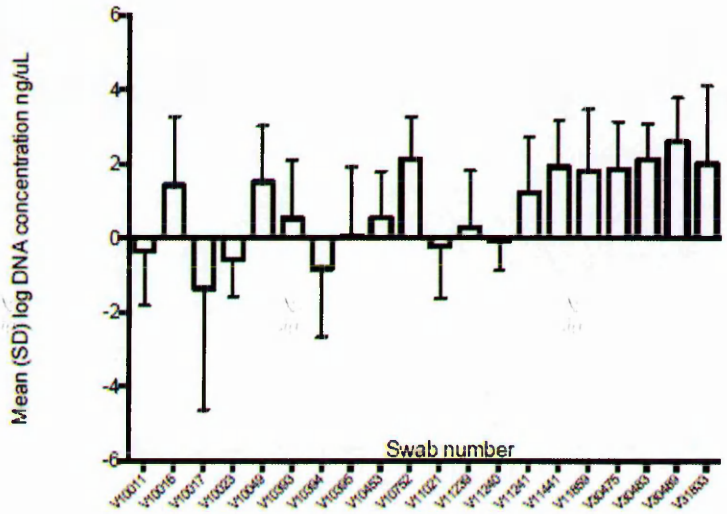
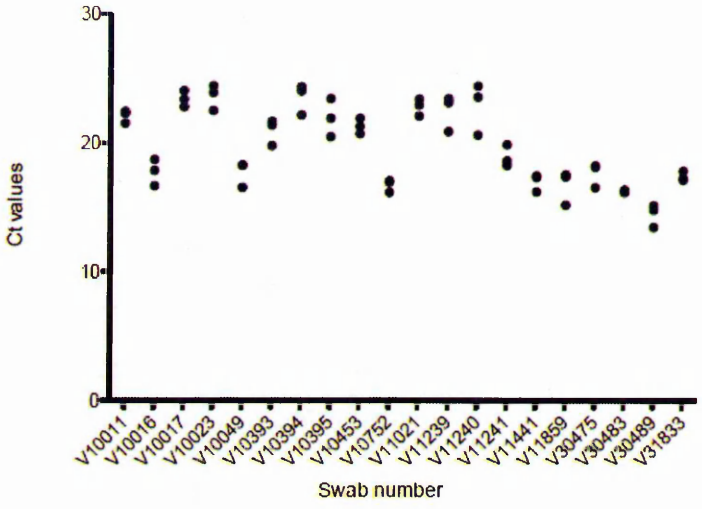
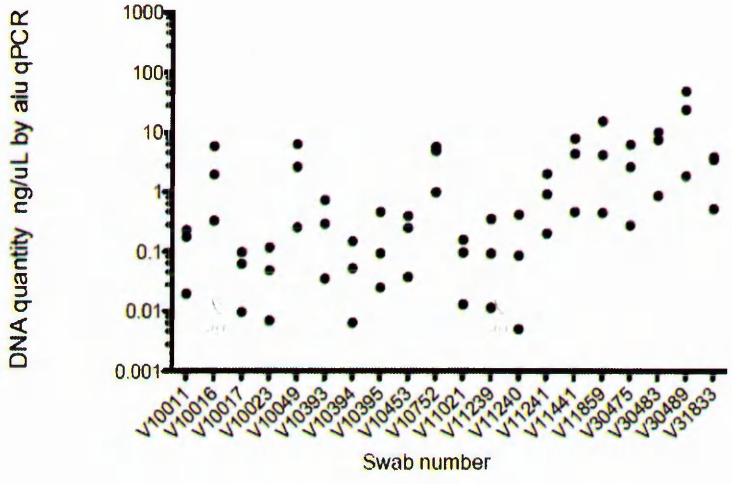


Figure 5.7: Reproducibility of *alu* qPCR for human DNA quantification in experiment three. Three extraction replicates of nasopharyngeal flocked swabs in viral transport medium. Experiment 3. **A:** *alu* qPCR Ct values. **B:** Human DNA concentration. **C.** Mean and standard deviation of *alu* Ct values. **D.** Mean and standard deviation of natural log human DNA concentration.

A Reproducibility of *alu* qPCR Ct values



B Human DNA concentration



C Mean and standard deviation of natural log human DNA concentration

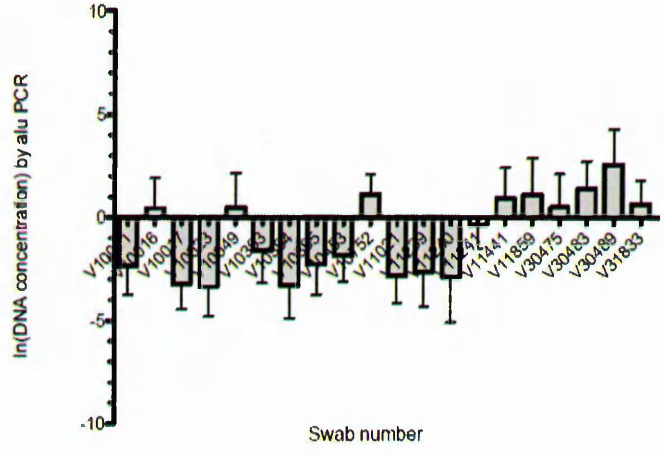


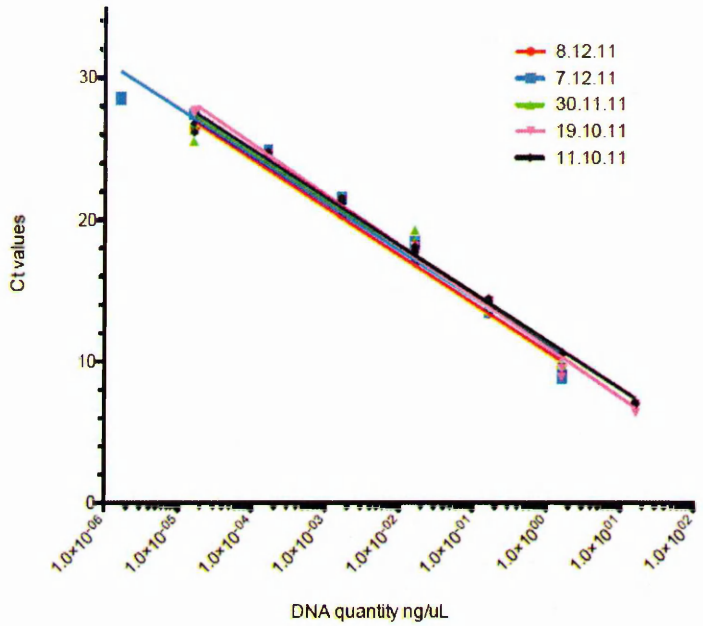
Figure 5.8: Reproducibility of *alu* qPCR for human DNA quantification in experiment four. Three extraction replicates of nasopharyngeal flocked swabs in viral transport medium. Experiment 4. **A:** *alu* qPCR Ct values. **B:** Human DNA concentration. **C:** Mean and standard deviation of natural log human DNA concentration.

The mean intra-swab standard deviation of natural log transformed human DNA concentration by *alu* qPCR in experiment three was 1.41ηg/μL, and in experiment four was 1.48ηg/μL. This compares with the inter-swab standard deviation in natural log human DNA concentration of 1.96ηg/μL, 2.02ηg/μL and 1.87ηg/μL for each set of replicates respectively.

The mean intra-swab standard deviation of natural log transformed pneumococcal concentration in *lytA* copies/μg human DNA was 1.25 copies/μg in experiment three and 1.14 copies/μg in experiment four. The inter-swab standard deviation in pneumococcal concentration in natural log *lytA* copies/μg human DNA was 2.30 copies/μg, 2.63 copies/μg and 2.35 copies/μg for each set of replicates respectively, in experiment four. The mean natural log pneumococcal concentration per μg of human DNA for each set of replicates was 14.93 copies/μg, 12.60 copies/μg and 12.99 copies/μg.

The standard curves for *lytA* qPCR from these reproducibility experiments are shown in chapter three, figure 3.6. For *alu* qPCR the standard curves from the reproducibility experiments are shown in figure 5.9.

A Five individual standard curves from *alu* qPCR



B Mean and standard deviation of Ct values from five *alu* qPCR standard curves

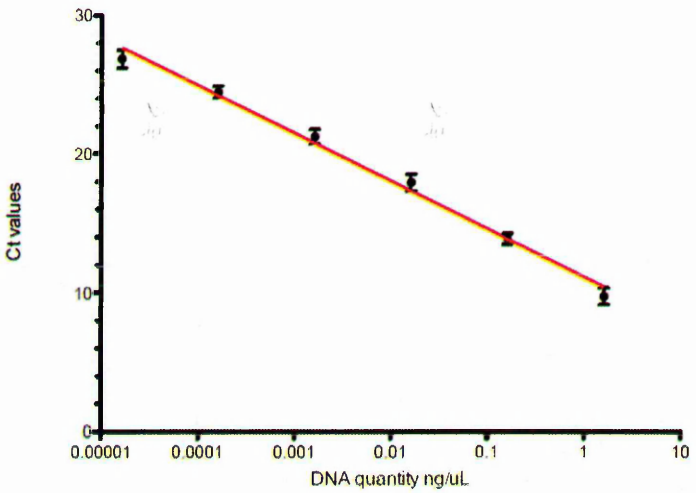
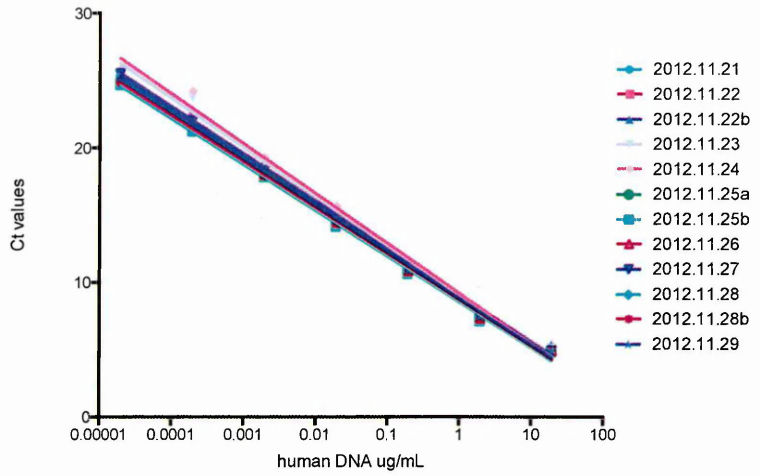


Figure 5.9: Variability in five standard curves from *alu* qPCR reproducibility experiments. A: Individual standard curves. **B:** Mean and standard deviation of Ct values for five *alu* qPCR standard curves.

The *lytA* qPCR standard curves from objective two, the study of pneumococcal concentration during episodes of viral upper respiratory tract infection, have been been shown in chapter three, figure 3.7. The *alu* qPCR standard curves from objective two can be seen in figure 5.10.

A Standard curves from 12 *alu* qPCR runs



B Mean and standard deviation of Ct values from 12 *alu* qPCR runs

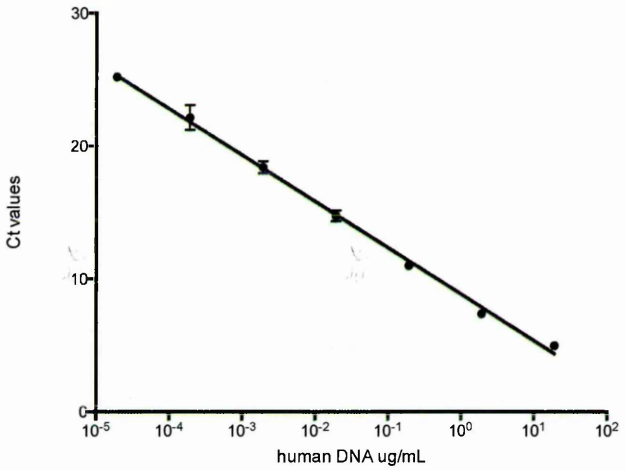


Figure 5.10: Standard curves for *alu* qPCR from nasopharyngeal swabs in viral transport medium from objective two. A: Standard curves from 12 *alu* qPCR runs. **B:** Mean and standard deviation of Ct values from 12 *alu* qPCR runs.

5.4.5.3 Conclusions

Pneumococcus is reproducibly detected by *lytA* qPCR from nasopharyngeal flocked swabs in VTM extracted by MagnaPure. Human DNA concentration by *alu* qPCR is more variable than pneumococcal concentration, and as this variation did not appear to come from the quantification standards it probably came from the dilution step, which had been shown to be necessary. Pneumococcal concentration adjusted for human DNA concentration had reasonable reproducibility, considering extraction replicates.

5.4.5.4 Next steps in human DNA quantification

Since the *alu* qPCR was not as reproducible as the single target copy *lytA* PCR, it was decided to try two other human DNA targets, with the *alu* qPCR to fall back on. Accordingly, a keratin gene target qPCR and an SFTPC (human surfactant protein C, NCBI accession number U02948) gene target qPCR were set up. The keratin gene target came from an NCBI BLAST search and primer BLAST was used to design the primers as well as the mfold website [191] to check the amplicon folding structure. For the SFTPC qPCR, the primers came from a published paper,[180] and the mastermix recipe and cycling conditions were adapted for locally available reagents and instruments. Both PCRs were successful, but the SFTPC PCR was pursued because it was working first. The SFTPC qPCR was demonstrated to be more reproducible than the *alu* qPCR and worked for nasopharyngeal samples in VTM and in STGG transport media. Unfortunately, when initially used for participant samples in the pilot study for objective two, from the household study of viral transmission, a half of the samples were completely negative for the SFTPC qPCR, presumably because the SFTPC qPCR is not as sensitive to low levels of human DNA as the multi-copy target *alu* qPCR is able to be. This line of work was

therefore dropped and the *alu* qPCR was used for human DNA quantification for the remainder of the project.

It was decided a priori, to examine objective two, nasopharyngeal pneumococcal concentration before, during and after viral URTI, both with and without adjustment for human DNA concentration by *alu* qPCR. Discussion of the effect of the *alu* qPCR adjustment can be found in chapters nine and twelve.

Chapter Six

6 Validation of laboratory methods for nasopharyngeal swab specimens in STGG

6.1 Introduction

Skim-milk-Tryptone-Glucose-Glycerol (STGG) is the WHO-recommended transport medium for bacterial culture of nasopharyngeal swabs for pneumococcal carriage.[168] This sample type was used in objectives one and three.

Sources of variation considered include the DNA extraction method, time during transport from the field (following the 2003 WHO guidelines[168]), and the volume of posterior nasopharyngeal secretions in the specimen. The serotype of pneumococcus was also considered as a potential source of variation, but it was beyond the scope of this investigation to examine.

6.2 The DNA extraction method

Potential DNA extraction methods were compared, using spiked samples and several different available extraction kits, as part of an intern's project, under the author's supervision. Manual methods rather than automated methods were considered, since we were not looking for a high-throughput method. The first objective was to choose the most appropriate kit for DNA extraction.

6.2.1 Methods

A suspension of the type strain of *S. pneumoniae*, ATCC 49619 was made in normal saline and adjusted to approximately 10^8 CFU/mL using a nephelometer (TREK Diagnostic Systems, Thermo Scientific, USA). STGG was then spiked with this pneumococcal suspension, to three different concentrations (10^7, 10^4 and 10^2 CFU/mL) to mimic patient specimens. Both colony counts and DNA extraction were performed on aliquots of the spiked specimens. Colony counting was performed by serial 1:5 dilution of the starting suspension (the simulated patient sample) with sterile STGG, then from each expected countable concentration dispensing in duplicate 100µL onto a pre-warmed 5% horse blood agar plate and spreading it using a wide-mouthed sterile wire loop, allowing the plate to absorb the suspension, then turning it upside down and placing in a 5% CO2 incubator at 35°C for 18-24 hours. Colony counting was performed manually the following day from all countable plates and a mean concentration for the starting suspension calculated.

DNA extraction was performed from 1mL aliquots of the starting suspension using four different extraction kits (ZR Fungal/Bacterial DNA Miniprep, Zymo, California, USA; Quick-gDNA Miniprep, Zymo, California, USA; Agilent gDNA kit, Agilent, California, USA; QIAamp DNA minikit, Qiagen, Germany) according to the manufacturer's instructions, prior to *lytA* qPCR. The experiment was repeated once for the gDNA Zymo kit and twice for the Qiagen kit, to check for reproducibility (one repeat sample was spiked at 10^6 cells/mL instead of 10^7 cells/mL). The *lytA* qPCR method can be seen in section 7.10 of chapter seven, and was performed according to *lytA* qPCR method for nasopharyngeal swab specimens, but with standards made from DNA extracted from *S. pneumoniae* ATCC 49619, measured against *lytA* plasmid standards (Fast-Track Diagnostics, Luxembourg).

6.2.2 Results

The pneumococcal concentrations by colony counting and the pneumococcal concentrations by quantitative PCR, for each concentration and extraction method studied, can be seen in table 6.1.

Table 6.1: Pneumococcal concentration from four different DNA extraction kits over a range of expected concentrations, by *lytA* qPCR and by colony-counting, from pneumococcus-spiked STGG. STGG = skim-milk tryptone glucose glycerol

Extraction method and expected concentration in spiked STGG, cells/mL	PCR quantity *lytA* copies/mL	Colony count quantity CFU/mL
Qiagen mini 10e6	133,080	1,562,500
Qiagen mini 10e4	2,606	1,562
Qiagen mini 10e2	89	16
Qiagen mini 10e7	2,975,580	11,718,750
Qiagen mini 10e4	5,236	11,719
Qiagen mini 10e2	53	117
Qiagen mini 10e7	1,556,000	38,650,000
Qiagen mini 10e4	2,480	38,650
Qiagen mini 10e2	32	387
Bact/Fung Zymo 10e7	44,900	60,000,000
Bact/Fung Zymo 10e4	71	60,000
Bact/Fung Zymo 10e2	negative	600
Agilent 10e7	343,333	3,030,000
Agilent 10e4	353	3,030
Agilent 10e2	18	30
gDNA Zymo 10e7	222,500	5,450,000
gDNA Zymo 10e4	515	5,450
gDNA Zymo 10e2	negative	55
gDNA Zymo 10e7	157,750	65,450,000
gDNA Zymo 10e4	272	65,450
gDNA Zymo 10e2	negative	655

6.2.3 Conclusions:

The Zymo kits (Zymo, California, USA) failed to extract pneumococcal DNA at the lowest concentrations, and the most consistent results were obtained for the QIAamp DNA mini kit extractions (Qiagen, Germany).

6.3 Time during transport of specimen after collection

The length of time a nasopharyneal swab in STGG spent traveling back from the field after collection, before reaching the laboratory for culture or DNA extraction, might affect quantitative results. Different serotypes of pneumococcus might have different survival characteristics in the nasopharyngeal swab STGG transport medium. We aimed to detect loss of viability of pneumococcus in STGG transport media over eight hours and loss of detectable pneumococci in STGG transport media by qPCR over eight hours. Although not powered to detect individual differences by serotype, each of the serotypes present in the 10-valent pneumococcal conjugate vaccine was included.

6.3.1 Methods:

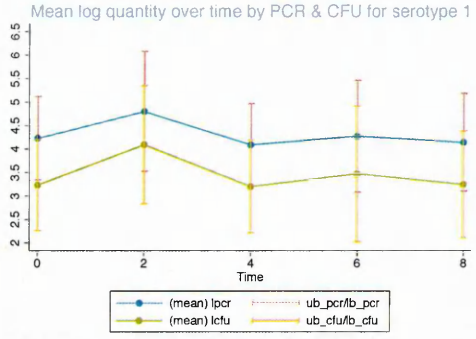
A QC Type Strain, *S. pneumoniae* ATCC 49619, and 30 clinical pneumococcal isolates, three of each PCV-10 serotype (i.e. 1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, 23F) were spiked into vials of STGG at five concentrations (10^6-10^2 cells/mL) to mimic nasopharyngeal swab inoculation. A nephelometer (TREK Diagnostic Systems, Thermo Scientific, USA) was used to measure the density of the pneumococcal suspensions prior to spiking, at 0.5 McFarland. Spiked STGG was held in a cool-box with an ice-pack in a tropical climate for eight hours, to mimic specimen transport from the field, as per the WHO

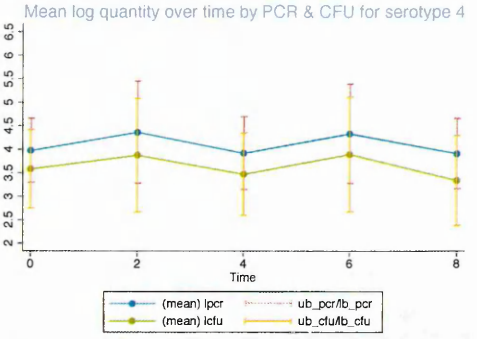
recommendations.[168] Aliquots were taken from each vial at 0, 2, 4, 6 and 8 hours for culture, and for DNA extraction (QiaAmp DNA minikit, Qiagen, Germany) for qPCR. Cultures were performed on 100µL of 1:5 serial dilutions in STGG of each aliquot in duplicate on 5% horse blood agar in 5% CO_2 at 35°C for 18-24 hours, and the mean colony count from all countable plates was taken for analysis. DNA extraction was performed from 1mL aliquots from each vial, otherwise according to the description of DNA extraction using the QIAamp manual method, from nasopharyngeal swabs in STGG, in section 7.6.2 in chapter seven.

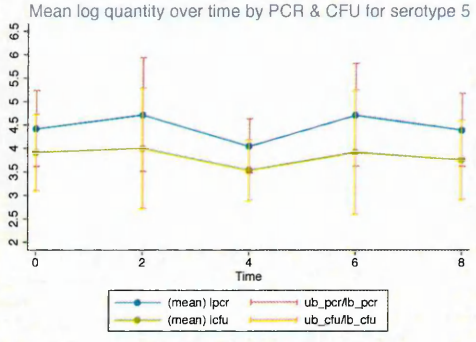
lytA qPCR was performed as described in *lytA* qPCR for nasopharyngeal swab specimens in section 7.10 in chapter seven, except for the final extraction:elution ratio adjustment, and with in-house standards. Standards for *lytA* qPCR were made by extracting DNA (QiaAmp DNA minikit, Qiagen, Germany) from a dense suspension of pneumococcus, performing colony counting on the original suspension and making a 1:10 serial dilution of the extracted DNA in molecular grade water. Subsequently, the standards were calibrated against commercially available *lytA* plasmid standards (Fast-Track Diagnostics, Luxembourg). Data points with <1000 copies/mL by qPCR were dropped from the analysis as this was considered to be below the limit of linearity for the qPCR. Linear regression of $\log_{10}$-transformed mean counts at two hour intervals in CFU/mL from culture and copies/mL from qPCR was performed to look for trends over time. A Wilcoxon signed-rank test was used to assess concentration by the culture method in CFU/mL compared to qPCR in copies/mL.

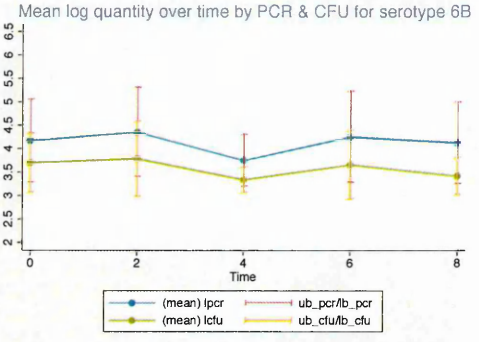
6.3.2 Results

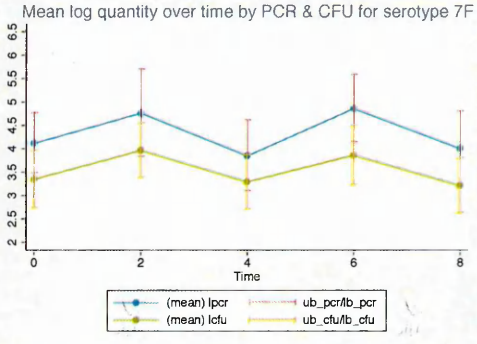
The mean $\log_{10}$ concentration of 10 serotypes of pneumococcus recovered from STGG by culture and colony counting in CFU/mL and by *lytA* qPCR in genome-copies/mL over eight hours under swab transport conditions can be seen in figure 6.1.

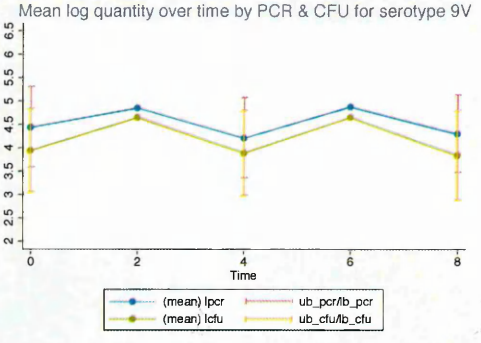

Mean log quantity over time by PCR & CFU for serotype 1
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu

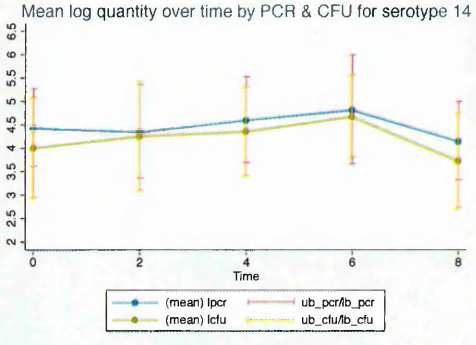

Mean log quantity over time by PCR & CFU for serotype 4
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu

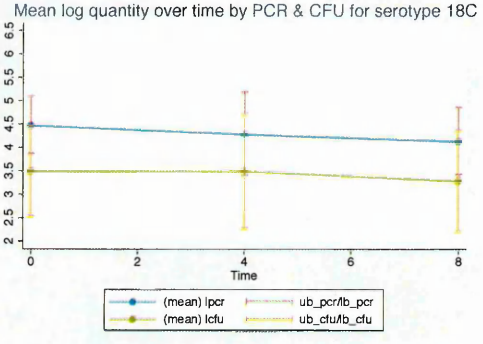

Mean log quantity over time by PCR & CFU for serotype 5
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu

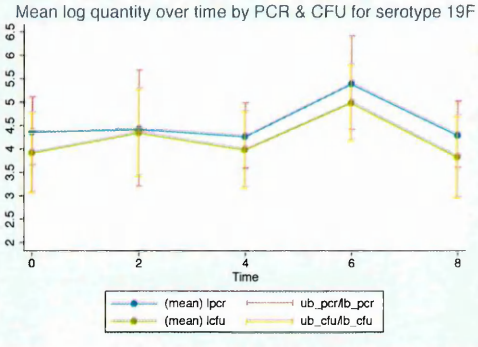

Mean log quantity over time by PCR & CFU for serotype 6B
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu

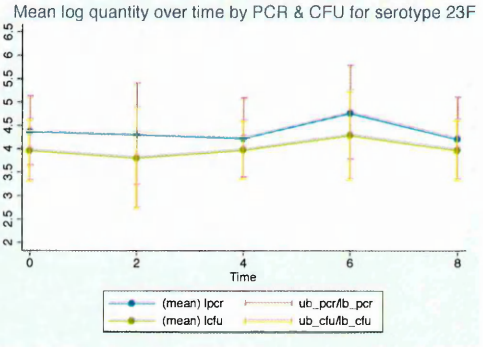

Mean log quantity over time by PCR & CFU for serotype 7F
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu


Mean log quantity over time by PCR & CFU for serotype 9V
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu


Mean log quantity over time by PCR & CFU for serotype 14
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu


Mean log quantity over time by PCR & CFU for serotype 18C
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu


Mean log quantity over time by PCR & CFU for serotype 19F
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu


Mean log quantity over time by PCR & CFU for serotype 23F
Time
(mean) lpcr
(mean) lcfu
ub_pcr/lb_pcr
ub_cfu/lb_cfu

Mean log_{10} concentration in genome-copies/mL by PCR (95% confidence intervals)

Mean log_{10} concentration in CFU/mL by colony counting (95% confidence intervals)

Figure 6.1: Mean log_{10} concentration of pneumococcus recovered from STGG by culture and colony counting in CFU/mL and by *lytA* qPCR in genome-copies/mL, over 8 hours under swab transport conditions, by serotype. Time on the x-axes is in hours.

The overall mean log concentration of pneumococcus recovered from STGG by culture and colony counting in CFU/mL and by lytA qPCR in genome-copies/mL, over eight hours under swab transport conditions, can be seen in figure 6.2.

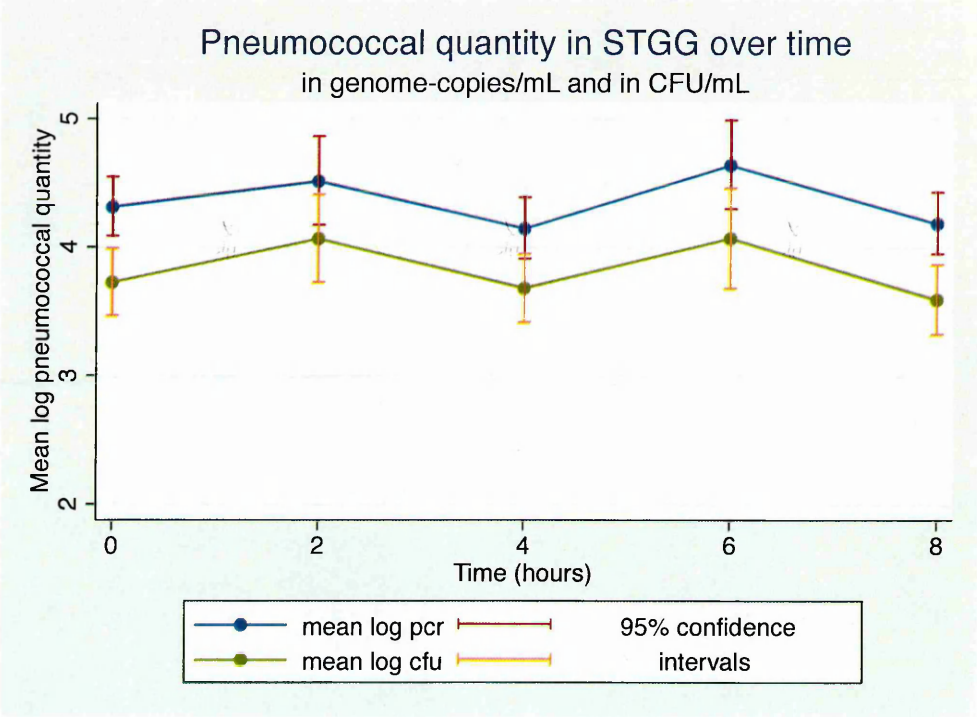


Figure 6.2: Mean log_{10} concentration of total pneumococcus recovered from STGG by culture and colony counting in CFU/mL and by *lytA* qPCR in genome-copies/mL, over 8 hours under swab transport conditions.

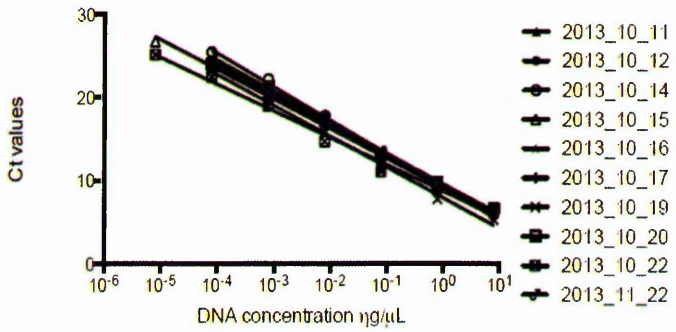
6.3.3 Conclusions

Over eight hours there was no demonstrable loss in detection of pneumococcus from STGG by either colony counting or by quantitative *lytA* PCR over ten invasive serotypes.

6.4 The volume of nasopharyngeal secretions in the STGG specimen

Following on from the work on human DNA quantification from the nasopharyngeal swab in viral transport medium, the same principles were considered for nasopharyngeal swabs in STGG. The mean and standard deviation of Ct values for standards in *alu* qPCR over ten runs performed on different days can be seen in figure 6.3.

A Ct values from *alu* qPCR standard curves



B Mean and standard deviation of Ct values from *alu* qPCR standard curves

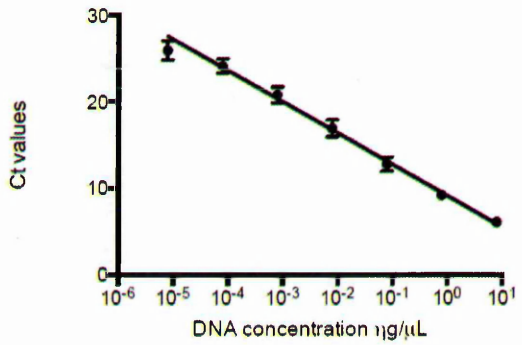


Figure 6.3: Standard curves for *alu* qPCR from nasopharyngeal swabs in STGG. A: Ct values for standard curves from 10 individual runs of *alu* qPCR. **B:** Mean and standard deviation of Ct values for standard curves from 10 individual runs of *alu* qPCR.

The specimen quality, sample dilution and reproducibility experiments described in chapter five were considered sufficient and not repeated for nasopharyngeal swabs in STGG.

It was decided a priori, to examine objectives one and three, the study of nasopharyngeal pneumococcal concentration in pneumonia and in health and the study of pneumococcal concentration after PCV-10 vaccination, using nasopharyngeal swabs in STGG, both with and without adjustment for human DNA concentration by *alu* qPCR. Discussion of the effects of *alu* qPCR on results and conclusions from objectives one and three, can be found in chapters eight, ten and twelve.

Chapter Seven

7 General Laboratory Methods

7.1 Pneumococcal culture

From stored frozen pure isolates: *Streptococcus pneumoniae* was cultured on 5% horse blood agar for 16-24 hours in 5% CO_2 at 35-37°C.

From nasopharyngeal swabs in STGG by the WHO standard method:[168] STGG was vortexed at 2500rpm for 20 seconds. 10µL of sample was aspirated with a pipette and dispensed onto half a 5% horse blood agar with 2.5µg/mL gentamicin plate and streaked with a sterile wire loop. Plates were incubated for 18-24 hours in 5% CO_2 at 35-37°C. Suspected pneumococcal colonies were subcultured to a 5% horse blood agar plate with an optochin disk and incubated for 18-24 hours in 5% CO_2 at 35-37°C. Isolates with a zone of inhibition of ≥14mm to optochin were considered to be *S. pneumoniae*. Isolates with a zone of inhibition of 9-13mm to optochin underwent a bile solubility test and were considered to be *S. pneumoniae* if they were bile soluble.

From nasopharyngeal swabs in STGG by the broth enrichment method:[165] STGG was thawed and vortexed for 10-20 seconds. 200µL of STGG was transferred into 5mL of Todd Hewitt broth (Becton Dickinson, MD, USA) supplemented with 5mg/mL yeast extract (Becton Dickinson, MD, USA) and 1mL rabbit serum (TCS Biosciences, Buckingham, UK). The inoculated broth was vortexed and incubated at 35-37°C for 4 hours in 5% CO_2. The broth was again vortexed and 10µL inoculated onto 5% horse blood agar plates with 2.5µg/mL gentamicin. Plates were incubated for 18-24 hours in 5% CO_2 at 35-37°C. *S. pneumoniae* was identified as above.

Isolates of *S. pneumoniae* were serotyped by latex agglutination and the quellung reaction, with conventional multiplex PCR 'serotyping' used as an internal quality control.

7.2 Colony counting

Starting suspensions for colony counting were always harvested from fresh overnight growth of bacteria on 5% horse blood agar at 35-37°C in 5% CO_2. Colony counting was performed by serial 1:5 dilution of the starting suspension in either sterile normal saline, phosphate-buffered saline or sterile STGG, up to nine times, labelling tubes A-I. Depending on the McFarland standard of the starting suspension, assuming that 0.5 McFarland is 1.5 x 10^8 cells/mL (http://www.pro-lab.com/inserts/McFarland.pdf), the expected countable concentrations of 30-300 cells per 100µL were calculated. From dilutions E-I 100µL was dispensed in duplicate onto pre-warmed 5% horse blood agar plates and spread using a wide-mouthed sterile, cooled wire loop. The agar plates were allowed to absorb the suspension before inversion and placement in a 5% CO_2 incubator at 35-37°C for 16-24 hours. Colony counting was then performed manually from all countable plates and a mean concentration for the starting suspension calculated in CFU/mL.

7.3 Serum antibiotic bioassay

20µL of serum was dispensed onto a blank 6mm filter paper disk that had been placed on a Mueller-Hinton agar plate streaked evenly with a lawn of *Staphylococcus aureus* ATCC 25923 at 0.5 McFarland, and incubated for 18-24 hours at 35-37°C with the agar side up. Any zone of inhibition was considered a positive result and indicative of the presence of antibiotics in the subject's serum.

7.4 DNA extraction using MagNA Pure

Whole blood in EDTA tubes was stored at 2-8°C for up to a week before DNA extraction. Whole blood samples were extracted using the MagNA Pure LC instrument (Roche, Indianapolis, USA) with the MagNA Pure LC DNA Isolation kit III according to the manufacturer's instructions. DNA was eluted from 200µL of whole EDTA blood into 100µL of elution buffer. The MagNA pure LC DNA isolation Kit III is optimised for extraction of bacterial and fungal DNA. It lyses bacterial cells with lysis buffer and proteinase K, and purifies DNA based on magnetic bead technology. DNA binds to the silica surface of the magnetic beads in the chaotropic salt conditions of the high ionic strength binding buffer. A negative extraction control (molecular grade water) and a positive extraction control (supernatant from a BACTEC blood culture bottle positive for *Streptococcus pneumoniae* added to whole EDTA blood) were included in each run. Up to 32 samples, including controls, could be run at a time.

7.5 DNA extraction using QIAxtractor

Whole blood EDTA samples underwent extraction of DNA from 200µL of sample in the QIAxtractor (Qiagen, Germany) automated instrument using the QIAxtractor DX kit for DNA (reference number 950107 for DX reagents and 950037 for DX plasticware) and Q Protocol 'QXT Liquid DNA', according to the manufacturer's instructions. The QIAxtractor DNA procedure uses the specific binding of nucleic acids to silica membranes using the bind, wash, and elute steps similar to the manual QiaAmp chemistry, but is a vacuum system rather than a centrifugation-based system. After nucleic acids in lysates bind to the membrane, contaminants are washed away and purified nucleic acids are subsequently eluted from the membrane. A negative extraction control (molecular grade

water) was included in each run. Up to 96 samples, including controls, could be included in each run.

7.6 DNA extraction using the QIAamp manual method

7.6.1 From whole blood samples

Whole blood in EDTA tubes was stored at 2-8°C for up to 72 hours before DNA extraction. DNA was extracted from 200µL of whole blood EDTA using the QIAamp DNA blood minikit (Qiagen, Germany) according to the manufacturer's instructions in the QIAamp DNA blood mini kit handbook. Working surfaces and pipettes were cleaned with DNA-off and 70% ethanol. Any sample that was less than 200µL was noted so as to adjust the extraction:elution ratio later when calculating the original DNA concentration from the quantity obtained by qPCR. 200µL of buffer AL was placed into labelled 1.5mL eppendorf tubes and 200µL of whole blood sample added to each tube, with negative controls (molecular grade water) at the end. 20µL of proteinase K was added to the AL-blood mixture and vortexed to mix. Samples were then incubated in a heating block or water bath at 56°C for 10 minutes. 200µL of ethanol was added to each tube, then vortexed and briefly spun in a microfuge to remove droplets from the lid. The total solution was then carefully transferred using a pipette onto a QIAamp Mini spin column in a collecting tube, avoiding wetting the rim. These were centrifuged at 6000 xg (8000 RPM) for 1 minute at room temperature. Columns were removed and placed in a fresh collecting tube and the filtrate discarded. 500µL of buffer AW1 was added to each column and centrifuged at 10,000 x g (13,000 RPM) for 1 minute at room temperature. Again the column was removed to a fresh collection tube and the filtrate discarded. 500µL of buffer AW2 was then added to the column, centrifuged at 10,000 xg (13,000 RPM) for 3 minutes at room temperature, the column removed to a fresh collection tube and the filtrate discarded.

An additional dry spin at 10,000 xg (13,000 RPM) was then performed for 30 seconds. Carefully the column was transferred into a labelled sterile eppendorf tube with the cap removed and the collection tube discarded. 100µL of buffer AE was added onto the column membrane and incubated for 5 minutes at room temperature before centrifugation at 6000 xg (8000 rpm) for 1 minute. The columns were then discarded and the eluted DNA immediately transferred to screw-capped Apex tubes for storage at -80°C. Working surfaces and pipettes were cleaned with DNA-off and 70% ethanol after use.

7.6.2 From nasopharyngeal swabs in STGG

Nasopharyngeal swabs in STGG transport medium (NP-STGG) were stored at -80°C and thawed on ice for DNA extraction. DNA was extracted from 200µL of STGG using the QIAamp DNA minikit (Qiagen, Germany). Before beginning, the working surfaces, racks and pipettes were cleaned using DNA-off and 70% ethanol. The thawed STGG tubes were vortexed for 20 seconds for proper mixing and immediately 200ul of STGG was pipetted into a 1.5mL microcentrifuge tube. STGG aliquots were centrifuged for 10 minutes at 5000 xg (7500 rpm). Using a fine tipped pastette, the supernatant was carefully aspirated, without touching the pellets, and discarded. 180µL of Buffer ATL was added to the pellets and then 20µL of proteinase K. This was then vortexed until the pellet dislodged from the bottom of the microcentrifuge tube. Samples were placed in the thermomixer at 56°C for 1 hour while shaking at 1000 rpm, or were vortexed 3 times during 1 hour in a heat block at 56°C if the shaking thermomixer was not available. After a brief centrifugation to remove condensate from the lids, 200µL of Buffer AL was added and the sample pulse vortexed 30 times. Samples were incubated at 70°C for 10 minutes, and briefly centrifuged again. 200µL of 100% ethanol was added to each sample, pulse vortexed, and briefly centrifuged. Using a pipette with 1000µL barrier tips, the mixture was transferred

to a QIAamp Mini spin column in a collecting tube without wetting the rim. This was centrifuged at 6000 xg (8000 rpm) for 1 minute, the spin column placed in a new collection tube and the tube containing the filtrate discarded. 500µL of Buffer AW1 was added to each sample, centrifuged at 6000 xg (8000 rpm) for 1 minute, the spin column placed in a new collection tube and the tube containing the filtrate discarded. 500µL of buffer AW2 was added to each sample, centrifuged at 20,000 xg (14,000 rpm) for 3 minutes, the spin column placed in a sterile 1.5mL microcentrifuge tube with the cap removed and the collection tube containing the filtrate discarded. 60µL of Buffer AE was added to each sample and incubated at room temperature for 1 minute before centrifugation at 6000 xg for 1 minute. Eluates in the microcentrifuge tube were transferred to screw-capped 0.5mL Apex tubes labelled with barcode cryolabels using the extract identification number (unique identification for that extraction, as samples could be extracted more than once). Extracted DNA was stored at -80°C.

7.6.3 From bacterial suspensions for pneumococcal microarray

NP-STGG samples were vortexed for 10-20 seconds and 1:10 and 1:100 dilutions were made in brain heart infusion broth (Oxoid, UK). 50µL of neat STGG suspension, the 1:10 and 1:100 dilutions were pipetted onto individual COBA media plates (streptococcal selective agar, Oxoid, UK) and spread evenly over the agar surface using sterile L-shaped plastic spreaders. Plates were incubated overnight at 37°C in 5% CO_2. The plate with the best high-density of distinct colony growth was selected for processing, and the approximate number of alpha-haemolytic colonies recorded. One mL of sterile phosphate-buffered saline (PBS) was added to the plate and the surface of the agar scraped gently using a sterile L-shaped plastic spreader, to make a suspension of all the colonies present. A 1mL pipette was used to transfer the resulting suspension to a 1.5mL

eppendorf tube. The suspension was centrifuged at 13,000 rpm for 5 minutes to pellet the bacteria. A lysozyme solution in lysis buffer for pre-extraction lysis was freshly prepared on each day of use. This consisted of 2190µL of H_2O, 50µL of 1M Tris-HCl, 10µL of 500mM EDTA and 250µL of Triton, for a total of 2.5mL of lysis buffer. Lysozyme (Sigma-Aldrich, Switzerland) was then dissolved into the lysis buffer at a concentration of 20mg/mL. 180µL of lysozyme solution was added to each bacterial pellet and it was incubated at 37°C for 60 minutes. Following this pre-extraction lysis, the QIAamp DNA mini kit (Qiagen, Germany, catalogue number 51304) was used according to the manufacturer's instructions. 20µL of proteinase K and 200µL of Buffer AL was added to each sample, which was then mixed by vortexing and incubated at 56°C for 60 minutes. Four µL of 100mg/mL RNase A was added to each sample, mixed by vortexing, incubated at room temperature for 5 minutes and briefly centrifuged. Samples were then incubated at 70°C for 10 minutes and briefly centrifuged again. 200µL of ethanol was added to each, mixed by vortexing and briefly centrifuged. The entire sample was then transferred to a QIAamp spin column in a collection tube, without wetting the rim, and centrifuged at 8000 rpm for 1 minute. The filtrate was discarded and the spin column placed in a clean collection tube. 500µL of buffer AW1 was added to the spin column and it was centrifuged at 8000 rpm for 1 minute. Again the filtrate was discarded and the spin column placed in a clean collection tube. 500µL of buffer AW2 was added to each to spin column and centrifuged at 13,000 rpm for 3 minutes. The filtrate was discarded and the spin column placed in a clean 1.5mL eppendorf tube. This was centrifuged at 13,000 rpm for 1 minute to remove any residual ethanol. The spin column was placed in a clean 1.5mL eppendorf tube. 100µL of molecular grade H_2O was pipetted on to the spin column membrane and allowed to stand for 1 minute at room temperature. This 100µL elution volume was reduced for plates with low colony counts, down as far as 20µL for samples

with <10 alpha-haemolytic colonies. Elution was performed by centrifugation at 13,000 rpm for 1 minute. The eluate containing extracted bacterial DNA was stored at 2-8°C for up to 24 hours until the DNA concentration could be measured using the Qubit Fluorometer (Life Diagnostics, California, USA) with the Qubit dsDNA high-sensitivity assay kit (catalogue number Q32851) and then stored at -80°C until use.

7.6.4 Buffy coat preparation for genomic DNA standards using the QIAamp DNA blood mini kit

Two vials of 10mL each of whole blood collected in EDTA were spun at 2500xg for 10 minutes at room temperature. Using a sterile Pasteur pipette, the buffy coat was collected from the middle layer of blood between the clear fraction and the erythrocytes from each tube and transferred into a clean 1.5mL eppendorf tube, obtaining approximately 1600µL of buffy coat. Twenty µL of proteinase K and 200µL of buffy coat extract were placed in each of eight sterile eppendorf tubes. Four µL of RNAse A was added to each tube and vortexed to mix before incubation at 56°C for 10 minutes. Tubes were briefly centrifuged and then 200µL of ethanol added to each. Again tubes were vortexed to mix and centrifuged before opening. The contents of each tube were carefully transferred into QIAamp mini spin columns and centrifuged at 8000 rpm for one minute before transferring the columns into clean collection tubes. Next 500µL of wash buffer AW1 was added and centrifugation for one minute at 8000 rpm was repeated. Then 500µL of wash buffer AW2 was added and centrifugation at 13,000 rpm was performed for three minutes. Fresh collection tubes were applied and a dry spin was performed for 30 seconds at 13,000 rpm before transferring the columns to sterile 1.5mL eppendorf tubes. Elution buffer at 100µL was added to each column and incubated at room temperature for two minutes. The columns in eppendorf tubes were then

centrifuged at 8000 rpm for one minute to collect the eluted DNA. DNA from all 8 tubes was combined, vortexed to mix and the concentration determined using a Qubit fluorometer (Life Technologies, California, USA), before storing at -20°C.

7.7 DNA extraction using easyMAG

The NucliSens easyMAG instrument (Bio-Merieux, France) uses silica extraction technology for automated isolation of nucleic acids. The sample is pre-processed to remove PCR inhibitors and release nucleic acid for isolation, diluted in lysis buffer to further remove any PCR inhibition effects and then isolation of total nucleic acid (TNA) is performed on the easyMAG instrument. TNA binds to silica particles in the presence of a high salt concentration. The silica particles have an iron core. A magnet is then used to extract silica particles with bound nucleic acid from solution containing organic sample matrix. Once the TNA has been removed from solution it is then purified by washing and eluted for downstream analysis.

Whole blood EDTA sample tubes were inverted gently several times to ensure the sample was homogeneous without vortexing. 200µL of whole blood sample was then immediately added to the respective labelled eppendorf tubes containing 800µL of Extraction Buffer 3. The entire 1mL of whole blood-Extraction Buffer 3 mixture was then added to the NucliSENS 2mL lysis buffer tube and vortexed immediately, holding the tube midway down the shaft to prevent sample splashing onto the tube cap. 140µL of vortexed, undiluted silica was then added to the sample-lysis buffer mix and vortexed immediately. The lysed samples were transferred from the NucliSENS lysis buffer 2mL tubes (total volume now 3mL) to the sample strips using a Pasteur pipette, without causing foam or droplets. A positive (a suspension of *S. pneumoniae* ATCC 49619 at approximately 10^3 cells/mL in whole blood) and negative (molecular grade water)

extraction control were included at the end of each run. The sample strips were then run on the NucliSENS easyMAG using the following parameters: Matrix = Whole blood, Protocol = Specific B 2.0.1, Volume = 1 mL (this refers to the 200µL of whole blood plus 800µL of Extraction Buffer 3 from the pre-processing step), Elute = 100µL, Type = Lysed, Priority = Normal, Lot number = lot of the 2mL lysis buffer. Extracted nucleic acids were removed for storage within 30 minutes of the run ending, and stored at -80°C.

7.8 DNA extraction from isolates using the boilate method

DNA extraction directly from bacterial isolates of *Streptococcus pneumoniae* was performed by the simple boilate method. A loopful of pure overnight growth from a blood agar plate was suspended in 60µL of Tris-EDTA (TE) buffer in an eppendorf tube, capped and placed into a heat block set at 100°C for 10 minutes. Each sample was then transferred to the centrifuge for a 10 minute spin at 14,000 rpm. Without disturbing the pellet, 50µL of supernatant containing bacterial DNA was pipetted with a barrier tip into a labelled cryovial for storage at -20°C until use.

7.9 *lytA* qPCR on whole blood specimens

lytA qPCR: Forward primer (F373): 5'-ACGCAATCTAGCAGATGAAGCA-3', Reverse primer (R424):5'-TCGTGCGTTTTAATTCCAGCT-3', Probe (Pb400): 5'-TGCCGAAAACGCTTGATACAGGGAG-3' (5' FAM; 3' MGB).[125] Mastermix containing 12.5µL of Gene Expression Mastermix (Applied Biosystems, Life Technologies, California, USA), 0.5µL of each of the forward and reverse primers and probe, 1µL of molecular grade water and 10µL of template DNA per reaction. Quantification standards consisting of *lytA* plasmids (Fast-Track Diagnostics, Luxembourg) diluted 1:10 from 10^7 copies/mL to 10^2 copies/mL were run in triplicate on every plate. Negative controls consisting of molecular

grade water were likewise run in triplicate. Cycling conditions of 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min were applied on an ABI-7500 real-time PCR instrument (Applied Biosystems, Life Technologies, California, USA). Exponential amplification curves with a cycle threshold (Ct) value of <40 cycles were considered positive and quantified using the standard curve. Standard curves had to have a slope between -3.1 and -3.6, for a PCR efficiency of at least 89%, and a correlation of >0.9. This was often achieved by omitting the 10^2 plasmid standard from the calculation, as discussed in chapter three, section 3.6. Quantified results were multiplied by a factor of 0.5, because the extraction was performed from 200µL of original sample and eluted into 100µL to be used for PCR. Data points with detected pneumococcal concentration below the lower limit of linearity and lower limit of detection of the assay were retained in the data set, with the understanding that accuracy and precision are affected in that range.

7.10 *lytA* qPCR on nasopharyngeal swab specimens

lytA qPCR: Forward primer (F373): 5'-ACGCAATCTAGCAGATGAAGCA-3', Reverse primer (R424):5'-TCGTGCGTTTTAATTCCAGCT-3', Probe (Pb400): 5'-TGCCGAAAACGCTTGATACAGGGAG-3' (5' FAM; 3' MGB). [125] Mastermix containing 12.5µL of Gene Expression Mastermix (Applied Biosystems, Life Technologies, California, USA), 0.5µL of each of the forward and reverse primers and probe, 6µL of molecular grade water and 5µL of template DNA per reaction. Quantification standards consisting of *lytA* plasmids (Fast-Track Diagnostics, Luxembourg) diluted 1:10 from 10^8 copies/mL to 10^2 copies/mL were run in triplicate on every plate. Negative controls consisting of molecular grade water were likewise run in triplicate. Cycling conditions of 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min were applied on an ABI-7500 real-time PCR instrument (Applied Biosystems, Life Technologies, California, USA). Exponential

amplification curves with a cycle threshold (Ct) value of <40 cycles were considered positive and quantified using the standard curve. Standard curves had to have a slope between -3.0 and -3.7, for a PCR efficiency of at least 86%, and a correlation of >0.9. This was often achieved by omitting the 10^2 plasmid standard from the calculation, as discussed in chapter three, section 3.6. Quantified results were multiplied by a factor of 60/200, because the extraction was performed from 200µL of original sample and eluted into 60µL to be used for PCR. Data points with detected pneumococcal concentration below the lower limit of linearity and lower limit of detection of the assay were retained in the data set, with the understanding that accuracy and precision are affected in that range.

7.11 *lytA* PCR on bacterial isolates

lytA PCR: Forward primer (F373): 5'-ACGCAATCTAGCAGATGAAGCA-3', Reverse primer (R424):5'-TCGTGCGTTTTAATTCCAGCT-3', Probe (Pb400): 5'-TGCCGAAAACGCTTGATACAGGGAG-3' (5' FAM; 3' MGB). [125] Mastermix containing 12.5µL of Gene Expression Mastermix (Applied Biosystems, Life Technologies, California, USA), 0.5µL of each of the forward and reverse primers and probe, 8.5µL of molecular grade water and 2.5µL of template DNA per reaction. A positive control consisting of *lytA* plasmids (Fast-Track Diagnostics, Luxembourg) at 10^4 copies/mL, or a boilate made from *Streptococcus pneumoniae* ATCC 49619, was included on every plate. Negative controls consisting of molecular grade water were run in triplicate. Cycling conditions of 50°C for 2 minutes, 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute were applied on an ABI-7500 real-time PCR instrument (Applied Biosystems, Life Technologies, California, USA). Exponential amplification curves with an exponential multicomponent rise were considered positive.

7.12 *alu* qPCR

alu qPCR: Forward primer; CAACATAGTGAAACCCCGTCTCT, Reverse primer; GCCTCAGCCTCCCGAGTAG. [190] Mastermix containing 12.5µL SYBR green (Applied Biosystems, Life Technologies, California, USA), 1.5µL forward primer, 1.5µL reverse primer, 4.5µL molecular grade water and 5µL of template DNA per reaction. Standards for quantification were made from total human DNA, extracted from a whole blood buffy coat preparation using QiaAmp DNA blood mini kit (Qiagen, Germany), the concentration measured using a Qubit fluorometer (Life Technologies, California, USA) and serial 1:10 dilutions made such that the highest standard used was 19ng/µL with six subsequent dilutions, all run in triplicate on every plate. Negative controls consisting of molecular grade water were likewise run in triplicate. Nucleic acid extracts were diluted 1:1000 in molecular grade water before being subjected to *alu* qPCR because *alu* is a multi-copy gene found thousands of times throughout the human genome. Cycling conditions were 50°C for 2 minutes, 95°C for 10 minutes, then 40 cycles of 95°C for 15 seconds, 59°C for 1 minute, followed by a slow ramp to a melt of 95°C for 30 seconds and 59°C for 15 seconds, on an ABI-7500 (Applied Biosystems, Life Technologies, California, USA). Melt curves were examined for all positive samples, with a broad peak at 77-79°C expected. The quantity of *alu* detected in the negative template controls was subtracted from each sample in the plate, before being multiplied by 1000 (to account for the dilution factor) and multiplied by 60/200 (to account for the extraction:elution ratio). Any samples with a quantity the same or less than the negative template controls, were repeated at a dilution of 1:100 instead of 1:1000.

7.13 Pneumococcal microarray

DNA extracts were stored at -80°C before shipping frozen to St George's University Hospital, London, UK, for microarray analysis.[192] Briefly, DNA was fluorescently labelled with ULS-Cy3 or ULS-Cy5 with the Genomic DNA ULS Labelling Kit (Agilent, California, USA). Fluorescently labelled samples were hybridized overnight to the BμG@S SP-CPSv1.4.0 microarray at 65°C, according to the Agilent array comparative genomic hybridization (CGH) protocol. The microarrays were then washed and scanned with an Agilent microarray scanner and intensity data analysed using feature extraction software. Intensity data was statistically analysed using a custom Bayesian hierarchical model to determine the serotype, or combination of serotypes, present in the sample and assign a relative abundance of each serotype detected. (Personal communication, Jason Hinds, St George's University Hospital, London, UK).

The serotype is determined by oligos concentrated on capsular locus (*cps*) genes; taking into account all capsular locus genes present. For discrimination of closely related serotypes, oligos were designed targeting the functional mutation causing the difference in polysaccharide structure, so that all 90 fully-sequenced serotypes could be distinguished.[16]

The genome backbone is determined by oligos targeting all genes in the core and accessory genomes of *S. pneumoniae* TIGR4 and R6, outside of the *cps* locus. The combination of serotype (from the cps locus) and microarray type (from the genome backbone) correlates with multi-locus sequence types (MLST) or whole-genome sequence-based clustering. Microarray type nomenclature does not apply across different serotypes e.g. type 2 serotype 23F does not have the same genome backbone as type 2 serotype 14. Serotype switches can be seen, such as vaccine escape mutants, where the

genome backbone does not change but the serotype does, as can acquisitions of a new strain with the same serotype. The genome backbone can be confounded by multiple serotype carriage, depending on the relative abundance and the types present. The genome backbone can also discriminate between non-pneumococcal streptococci and *S. pneumoniae.*

Chapter eight

8 Objective one: Nasopharyngeal pneumococcal concentration in different disease states

8.1 Introduction

Pneumonia is the leading cause of death in children less than five years of age both in sub-Saharan Africa and in the world [1] and pneumococcus is the leading cause of pneumonia.[2] Yet despite having the highest burden of disease, pneumococcal pneumonia remains difficult to diagnose in young children. Children do not expectorate sputum for sampling, lung aspirates are usually considered too invasive to perform and blood cultures are too insensitive. Urinary antigen testing is probably helpful in the diagnosis of community-acquired pneumonia in adults,[193] but nasopharyngeal pneumococcal carriage is thought to cause false-positives in urine antigen testing in children.[194] Culture of nasopharyngeal swabs reveals pneumococcus among the majority of children under the age of five years.[195] It has been shown that quantitative methods such as quantitative PCR can be helpful in making the diagnosis of pneumococcal pneumonia among adults in South Africa [126] and that children with pneumonia had higher nasopharyngeal pneumococcal concentration than healthy controls in Vietnam.[99] We thought it possible that quantitative nasopharyngeal pneumococcal concentration could be the key to understanding nasopharyngeal sampling in the diagnosis of pneumonia among children.

Pneumococcal disease is thought to occur soon after nasopharyngeal acquisition of a virulent strain of pneumococcus.[66] It is plausible, but not proven, that an increase in pneumococcal density in the nasopharynx by the virulent strain is the next step which

then leads to pneumonia or invasive disease. This putative increase in pneumococcal nasopharyngeal concentration could be potentiated by presence of a viral upper respiratory tract infection (URTI). It is known that viral infections are temporally associated with pneumonia,[71,73] that viral URTI may directly precede pneumococcal pneumonia,[75] that respiratory viruses increase nasopharyngeal pneumococcal density, transmissibility and invasiveness in mice[82,89,95] and increase bacterial adherence to respiratory cell lines *in vitro*.[77]

If the supposition that an increase in pneumococcal concentration in the nasopharynx is the step which then leads to pneumonia or invasive disease is correct, then nasopharyngeal pneumococcal concentration ought to be lowest among healthy children, increased among children with symptoms of URTI, and highest among children with pneumococcal pneumonia. If this is the case, then quantitative methods of sampling the nasopharynx could potentially be used as part of the paediatric diagnosis of pneumonia. A caveat to this is that while Albrich *et al* were able to set a threshold for nasopharyngeal pneumococcal concentration in the diagnosis of pneumonia in adults,[126] Vu *et al* were unable to do so in their study of childhood pneumonia.[99]

It was not practical to study a large cohort of well children through the continuum of good health, viral URTI, pneumonia and bacteraemic pneumonia, sampling their nasopharynx at each point along the way. Instead we aimed to describe the concentration of *S. pneumoniae* in different groups of children by using the *lytA* quantitative PCR assay on nasopharyngeal swab specimens from the following groups;

(i) healthy children recruited as community controls from a large study of childhood pneumonia.

(ii) children with symptoms of upper respiratory tract infection recruited as community controls from a large study of childhood pneumonia.

(iii) children with WHO-defined severe or very-severe pneumonia admitted to Kilifi County Hospital from a large study of childhood pneumonia.

It was assumed that children in the latter group, with severe or very severe pneumonia, would contain cases of pneumococcal pneumonia.

8.1.1 Hypotheses

Nasopharyngeal pneumococcal concentration is higher in children with WHO-defined severe or very severe pneumonia than in children in the community who have URTI symptoms. Nasopharyngeal pneumococcal concentration is lowest among children in the community who are well.

8.2 Methods

8.2.1 Sample selection

The samples for this study came from the Pneumonia Etiology Research for Child Health (PERCH) project. PERCH is a multi-site case-control study of the causes of pneumonia in children aged one month to five years, which took place in Kilifi, Kenya as well as in the Gambia, South Africa, Mali, Zambia, Thailand and Bangladesh over 2011-2014. In Kilifi, children residing in the Kilifi Health and Demographic Survey System (KHDSS) area admitted to Kilifi County Hospital with the syndromes of WHO-defined severe or very-severe pneumonia were enrolled from August 2011. Community controls who were well, or who had symptoms of acute URTI were age-matched monthly from the KHDSS.[196] The presence of URTI was defined as cough or runny nose or at least one of; ear discharge,

wheeze, difficulty breathing and at least one of fever or sore throat. In August 2012 after one year of enrolment, 70 community controls with URTI, the most constrained group, were randomly selected in four age strata; 1-5 months, 6-12 months, 13-24 months and 25-59 months. Seventy children from the group with WHO-defined severe or very severe pneumonia as well as 70 healthy children from the community control group were matched for month of enrolment and for age to the controls with URTI.

The selected children were stratified by age because younger children are more likely to carry pneumococcus [195] and were expected to have higher pneumococcal nasopharyngeal concentration.[103] Assuming a mean nasopharyngeal pneumococcal concentration of 7 $\log_{10}$ µg/mL in the group with pneumonia and 6 log µg/mL in either of the control groups, with a standard deviation of 1.5 $\log_{10}$ µg/mL in both groups,[99] and a sample size of 60 in each group, a two-sample comparison of means power calculation [197] yielded 95% power to detect the 1 log difference in pneumococcal concentration with a significance level of 0.05. Seventy participants were then selected from each group to allow for missing samples.

8.2.2 Laboratory methods

Nasopharyngeal swab samples in STGG collected at enrolment into the studies of pneumonia were taken from -80°C storage, DNA was extracted using the QIAamp DNA mini kit from Qiagen (a full description of the method can be found in section 7.6.2, chapter seven). *lytA* qPCR was performed to quantify pneumococcal load and *alu* qPCR to quantify the amount of human DNA present in the nasopharyngeal swab transport medium (sections 7.10 and 7.12 in chapter seven). As most swabs were *lytA* positive, it seemed possible that among the few cases where pneumococci were not detected that this might just as easily be attributable to the sensitivity of the assay as to an indication of

true negativity. Additionally, the lower limit of linearity of the assay was at the approximate lower limit of detection, 1000 copies/mL, so values <1000 copies/mL could be assumed to be inexact. Therefore all swabs with a raw pneumococcal concentration (unadjusted for extraction:elution ratio and for human DNA) of less than 1000 copies/mL, including negatives, were set to half this value, 500 copies/mL. Data on the lower limit of detection of the assay can be found in section 3.6 of chapter three. The pneumococcal concentration per µg of human DNA present in the transport medium was then calculated.

All children had whole blood samples at enrolment collected into EDTA, DNA extracted by either the QIAamp DNA mini kit from Qiagen, or by the EasyMag automated extraction instrument from BioMerieux (sections 7.6.1 and 7.7 in chapter seven). These then underwent *lytA* qPCR (section 7.9 in chapter seven) both for the PERCH study, and for objective four of this thesis.

As laboratory lead for the PERCH study in Kenya the author had access to additional data on the participants; for pneumonia cases this included presence of pneumococcal bacteremia by blood culture, presence of pneumococcus in culture of induced sputum, administration of antibiotic therapy prior to nasopharyngeal swab collection, and presence of radiologically-confirmed pneumonia. Data available for cases and controls included evidence of antibiotic use by serum bioassay (section 7.3 in chapter seven); growth of pneumococcus from NP-STGG by broth enrichment culture (section 7.1 in chapter seven); presence of 33 respiratory viruses and bacteria in a nasopharyngeal flocked swab and oropharyngeal swab combined in Copan universal transport medium (UTM), extracted using the EasyMag (BioMerieux, France) automated instrument, by multiplex real-time PCR (Fast-Track Diagnostics [FTD], Luxembourg). The FTD respiratory

multiplex PCR also included a *lytA* target. For this analysis, presence of *Mycoplasma pneumoniae, Chlamydophila pneumoniae* or *Bordetella pertussis* were included in the definition of respiratory-virus positive together with the respiratory viruses influenza A, B and C, parainfluenza viruses 1, 2, 3 and 4, human metapneumovirus, respiratory syncitial virus A and B, coronaviruses OC43, NL63, 229E and HKU1, bocavirus, adenovirus, cytomegalovirus, parechovirus/enterovirus and rhinovirus.

8.2.3 Analysis

Log transformation was performed to improve visualisation of the data, but did not completely normalise the data so non-parametric analysis of non-transformed data proceeded. A Kruskal-Wallis test was used to compare the median pneumococcal load between all three groups (pneumonia cases, URTI controls and well controls), and a two-sample Wilcoxon rank-sum test for each two-group comparison. The output was box-and-whisker plots of $\log_{10}$ pneumococcal load by symptomatic group.

The intention was to analyse the group with WHO-defined severe/very severe pneumonia in several strata including: presence or absence of respiratory virus in nasopharyngeal samples; finding of pneumococcus from bacterial culture of induced sputum; radiologically-confirmed pneumonia; and pneumococcal bacteraemia. The effect of these stratified or restricted analyses on the matching for age group and month of enrolment was checked by constructing tables of proportions in each group and by using Fisher's exact test. Presence of respiratory virus was assumed to have the potential to increase nasopharyngeal pneumococcal concentration in both the URTI and pneumonia groups. Presence of pneumococcus in bacterial culture of induced sputum, presence of radiologically-confirmed pneumonia and presence of pneumococcal bacteraemia were all assumed to concentrate pneumococcal aetiology cases among the WHO-defined

pneumonia cases, to improve the likelihood of seeing a difference between this group and the control groups.

The variation in pneumococcal concentration by group, age, season, presence of a respiratory virus, growth of pneumococcus in bacterial culture of induced sputum, radiologically-confirmed pneumonia, antibiotics given prior to sample collection and positivity of *lytA* PCR in whole blood was described using a Kruskal-Wallis or Wilcoxon rank-sum test depending on the number of categories of the dependent variable.

In addition a multivariable logistic regression, by backwards stepwise selection, was performed in order to adjust for age and month of enrolment in case of loss of matching in the stratified/restricted analyses. The likelihood ratio test was used to examine the variables in the final model for interactions.

The analysis comparing median pneumococcal concentration between the three symptomatic groups was repeated without adjustment for the concentration of human DNA present in the transport medium for comparison.

The proportion of pneumonia cases and of controls with URTI who had coryza documented was compared by chi-squared test.

Statistical analyses were performed in STATA version 11.

8.2.4 Ethical Clearance

The study was approved by the Kenyan National Ethical Review Committee (SSC 1932). In the original PERCH study consent was obtained from guardians of all children studied.

8.3 Results

The *lytA* qPCR assay was positive among 206/209 (98.6%) of swabs in STGG. By broth-enrichment culture 149/209 (71.3%) of nasopharyngeal swabs had grown pneumococcus; 51/60 (73.9%) of healthy controls, 62/70 (88.6%) of controls with URTI symptoms and 36/70 pneumonia cases (51.4%), p<0.001. None of the swabs that were *lytA* qPCR negative had grown pneumococcus, but 57 swabs that failed to grow pneumococcus were *lytA* qPCR-positive. Looking at the nasopharyngeal/oropharyngeal swab in UTM, the *lytA* target in the FTD respiratory multiplex PCR was positive in 168/208 (80.8%) samples. The distribution of $\log_{10}$-transformed pneumococcal concentration from the nasopharyngeal swab in STGG, after setting all *lytA* results <1000 copies/mL to a value of 500 copies/mL, adjusting for extraction:elution ratio and for human DNA concentration, can be seen in figure 8.1.

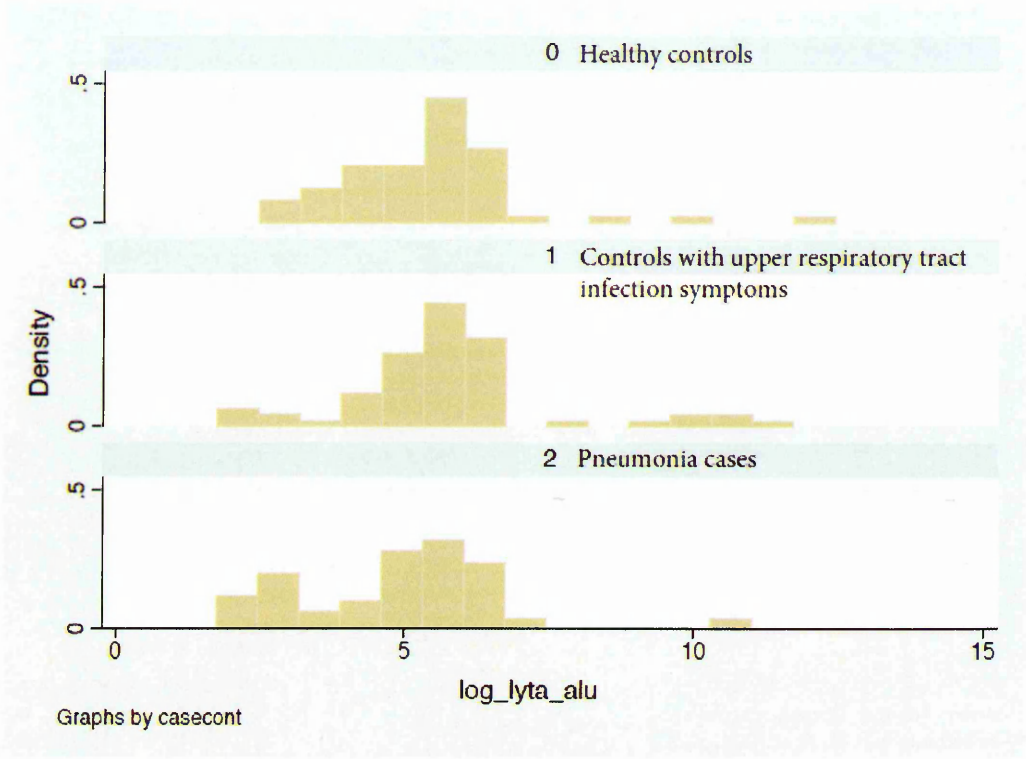


Figure 8.1: Distribution of nasopharyngeal pneumococcal concentration in log_{10} *lytA* copies/µg of human DNA by case-control status.

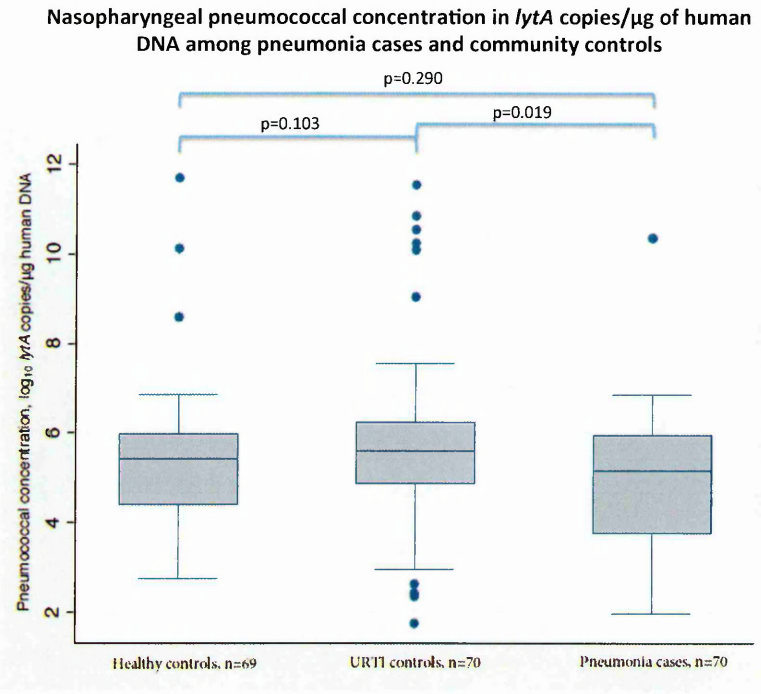
The variation in median pneumococcal concentration per µg of human DNA present in the transport medium in this entire sample set can be seen in table 8.1. The median nasopharyngeal pneumococcal concentration per µg of human DNA present in the transport medium was 148,148 copies/µg human DNA among pneumonia cases, 415,288 copies/µg human DNA among controls with URTI symptoms and 265,695 copies/µg human DNA among healthy controls. A Kruskal-Wallis test found a statistically significant difference between the three groups with a p-value of 0.046. Comparing each group with the other, by Wilcoxon rank sum test, only the difference between the cases and the controls with URTI symptoms was significant. This can be seen in figure 8.2.

Table 8.1: Variation in nasopharyngeal pneumococcal concentration by *lytA* qPCR by clinical characteristic

Characteristic	n	Median *lytA* concentration (copies/µg human DNA)	IQR		p-value*
Age group:					0.999
1-5 months	34	358,129	28,444	1,828,694	
6-12 months	51	298,899	41,854	1,001,807	
13-24 months	32	378,638	79,251	971,056	
25-59 months	21	265,695	15,473	1,709,503	
Month of the year:					0.114
January-March	78	169,243	17,587	985,799	
April-June	47	275,336	12,803	1,742,496	
July-September	41	597,780	76,986	1,944,094	
October-December	42	319,934	43,957	653,641	
Case/control status:					0.046
healthy controls	69	265,695	26,057	980,885	
controls with URTI	70	415,288	76,986	1,742,496	
cases of WHO-defined pneumonia	70	148,149	5,795	877,187	
Respiratory virus in nasopharynx:					0.108
Absence of respiratory virus	51	223,751	3,773	1,081,401	
Presence of respiratory virus	157	265,695	43,982	1,513,495	
Induced Sputum culture:					0.142
Absence of pneumococcus	48	91,608	1,408	625,711	
Presence of pneumococcus	18	460,490	126,181	1,960,489	
Chest X-ray:					0.552
No radiologically-confirmed pneumonia	41	126,181	16,815	653,641	
Radiologically-confirmed pneumonia	12	278,956	5,967	2,040,614	
Antibiotics:					0.203
No antibiotics given in hospital prior to sample collection and antibiotic bioassay negative	165	285,389	36,469	1,650,351	
Antibiotics given in hospital prior to sample collection or antibiotic bioassay positive	30	207,773	17,587	597,780	
lytA PCR from whole blood:					0.199
lytA PCR negative from whole blood	175	255,884	26,057	1,059,737	
lytA PCR positive from whole blood	15	398,715	117,685	3,630,154	

*Kruskal-Wallis where >2 categories; Wilcoxon rank-sum where 2 categories

A



B

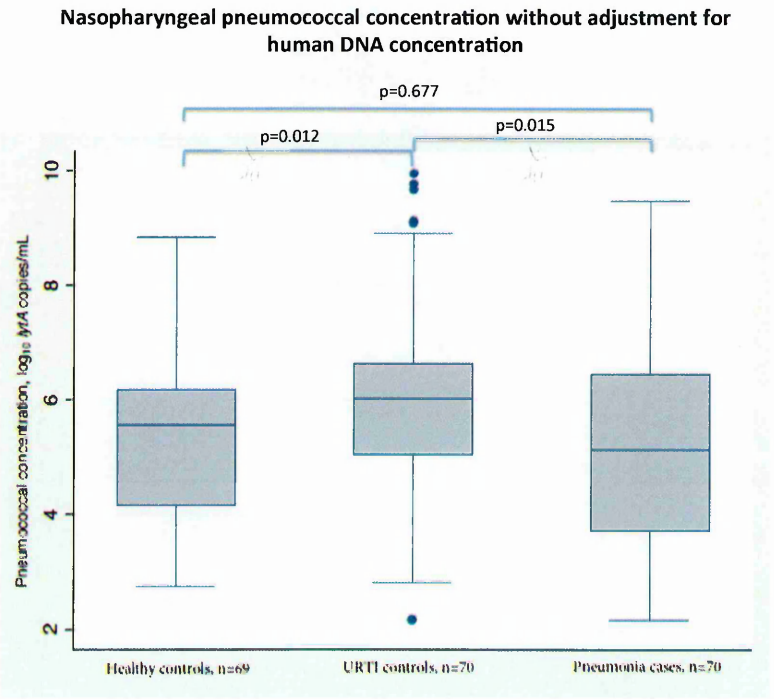


Figure 8.2: Nasopharyngeal pneumococcal concentration among healthy community controls, community controls with upper respiratory tract infection (URTI) symptoms,

and cases of WHO-defined severe or very severe pneumonia. P-values are by Wilcoxon rank-sum test. **A:** Pneumococcal concentration adjusted for the concentration of human DNA in the nasopharyngeal swab transport medium. **B:** Pneumococcal concentration without adjustment for human DNA concentration.

The pattern of the pneumococcal concentration across the three groups of healthy controls, controls with URTI and pneumonia cases did not change when stratified by presence of a respiratory virus in the nasopharynx. This can be seen in figure 8.3.

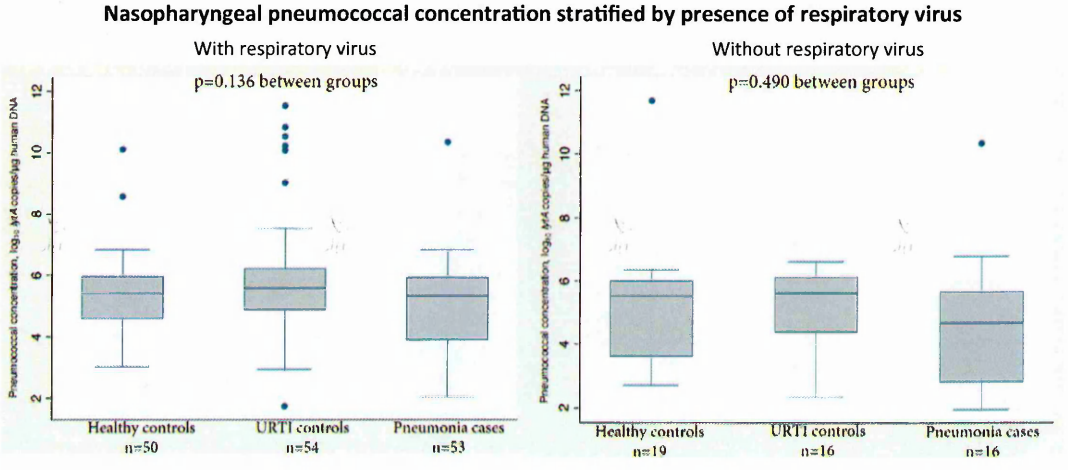


Figure 8.3: Nasopharyngeal pneumococcal concentration stratified by presence of a respiratory virus. Pneumococcal concentration in $\log_{10}$ *lytA* copies/µg human DNA among healthy community controls, community controls with upper respiratory tract infection (URTI) symptoms, and cases of WHO-defined severe or very severe pneumonia, stratified by presence of a respiratory virus by multiplex PCR of a combined nasopharyngeal and oropharyngeal swab specimen. P-values are by Kruskal-Wallis test.

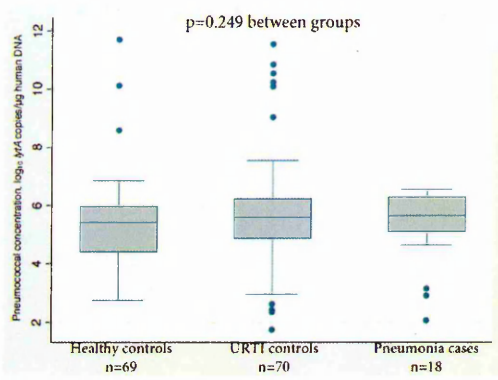
Cases from whom *S. pneumoniae* was grown in culture of induced sputum samples did not have a statistically significant difference in quantity of pneumococcus in the nasopharynx than controls. Controls did not have induced sputum collected. The distribution can be seen in figure 8.4.

Restricting cases to those with radiologically-confirmed pneumonia did not substantially alter the distribution of results, as can be seen in figure 8.4.

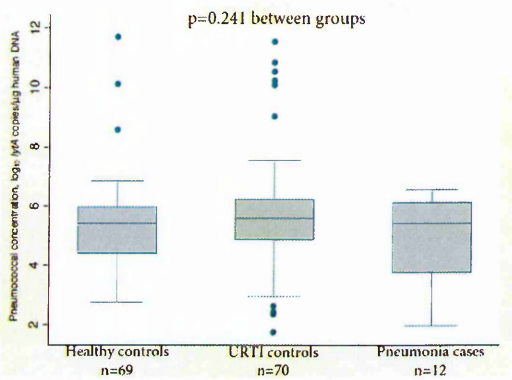
None of the cases had pneumococcal bacteraemia diagnosed by blood culture.

When pneumonia cases were restricted to only those who had not yet received antibiotic therapy in hospital and who had no evidence of antibiotic use on a serum antibiotic bioassay, and controls were restricted to those who had no evidence of antibiotic use on the bioassay, the differences in nasopharyngeal pneumococcal concentration between groups were reduced and the overall pattern was similar. This can be seen in figure 8.4. There was no statistically significant difference between groups overall (p=0.173), between healthy and URTI controls, between healthy controls and pneumonia cases, or between URTI controls and pneumonia cases.

A: *S. pneumoniae* in induced sputum



B: Radiologically-confirmed pneumonia



C: No antibiotics

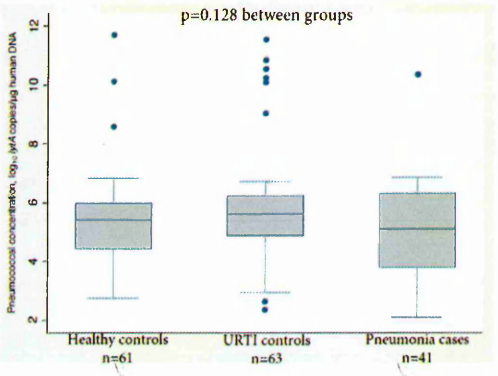


Figure 8.4: Nasopharyngeal pneumococcal concentration, restricted analyses.

Pneumococcal concentration in $\log_{10}$ *lytA* copies/µg human DNA among healthy community controls, community controls with upper respiratory tract infection (URTI) symptoms, and cases of WHO-defined severe or very severe pneumonia. **A:** Pneumonia cases were restricted to those with growth of *S. pneumoniae* on culture of induced sputum. **B:** Pneumonia cases were restricted to those with radiologically-confirmed pneumonia. **C:** Controls were restricted to those without evidence of antibiotic use by the serum antibiotic bioassay and cases of pneumonia were restricted to those without evidence of antibiotic use by the serum antibiotic bioassay and without antibiotics having been given in hospital prior to sample collection.

Repeating the analysis without adjusting the pneumococcal concentration for the amount of human DNA present in the specimen transport medium gave similar results. The median log nasopharyngeal pneumococcal concentration was 138,281 copies/mL among pneumonia cases, 1,079,054 copies/mL among controls with URTI symptoms and 376,047 copies/mL among healthy controls. The Kruskal-Wallis test yielded a p-value of 0.017 between the three groups. Comparing each group with the other, only the controls with URTI symptoms were significantly different from either the pneumonia cases or the healthy controls. This can be seen in figure 8.2.

The distribution of the proportions of each age group across healthy controls, controls with URTI and pneumonia cases was not significantly different; after stratification for presence of respiratory virus; on restriction to presence of pneumococcus on culture of induced sputum among cases; on restriction to radiologically-confirmed pneumonia among cases; and on restriction to cases and controls without evidence of antibiotic use. These can be seen in table 8.2.

Table 8.2: Effect of stratification and restriction on matching by age and month of enrolment. A: Stratification by presence of respiratory virus. **B:** Restriction to cases with culture of pneumococcus from induced sputum. **C:** Restriction to cases with radiologically-confirmed pneumonia. **D:** Restriction to cases and controls with no evidence of antibiotic use prior to sample collection.

A

No respiratory virus

Age group at enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
1-5 months	6	33.3	5	31.3	5	31.3
6-12 months	5	27.8	3	18.8	2	12.5
13-24 months	4	22.2	3	18.8	2	12.5
25-59 months	3	16.7	5	31.3	7	43.8
Fisher's exact p=0.762						
With respiratory virus						
1-5 months	11	22.0	12	22.2	12	22.6
6-12 months	20	40.0	23	42.6	23	43.4
13-24 months	12	24.0	13	24.1	14	26.4
25-59 months	7	14.0	6	11.1	4	7.6
Fisher's exact p=0.981						

No respiratory virus

Month of enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
Aug 2011	0	0	0	0	1	6.2
Sep 2011	2	11.1	1	6.3	0	0
Oct 2011	2	11.1	3	18.8	2	12.5
Nov 2011	3	16.7	2	12.5	3	18.8
Dec 2011	1	5.6	1	6.3	0	0
Jan 2012	0	0	1	6.3	4	25
Feb 2012	4	22.2	2	12.5	1	6.3
Mar 2012	3	16.7	2	12.5	2	12.5
Apr 2012	1	5.6	1	6.3	0	0
May 2012	0	0	0	0	1	6.3
Jun 2012	1	5.6	2	12.5	1	6.3
Jul 2012	1	5.6	0	0	1	6.3
Aug 2012	0	0	1	6.3	0	0
Fisher's exact p=0.850						
With respiratory virus						
Aug 2011	3	6	4	7.4	2	3.8
Sep 2011	0	0	1	1.9	2	3.8
Oct 2011	3	6	3	5.6	3	5.7
Nov 2011	4	8	4	7.4	4	7.6
Dec 2011	1	2	1	1.9	2	3.8
Jan 2012	9	18	8	14.8	5	9.4
Feb 2012	2	4	4	7.4	5	9.4
Mar 2012	8	16	9	16.7	9	17
Apr 2012	2	4	2	3.7	3	5.7
May 2012	5	10	6	11.1	5	5.7
Jun 2012	6	12	5	9.3	6	11.3
Jul 2012	5	10	6	11.1	5	9.4
Aug 2012	2	4	1	1.9	2	3.8
Fisher's exact p=1.000						

B

Cases restricted to those with induced sputum culture of pneumococcus

Age group at enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
1-5 months	17	25	17	24.3	4	22.2
6-12 months	25	36.8	26	37.1	8	44.4
13-24 months	16	23.5	16	22.9	4	22.2
25-59 months	10	14.7	11	15.7	2	11.1
Fisher's exact p=0.999						

Cases restricted to those with induced sputum culture of pneumococcus

Month of enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
Aug 2011	3	4.4	4	5.7	1	5.6
Sep 2011	2	2.9	2	2.9	0	0
Oct 2011	5	7.4	6	8.6	1	5.6
Nov 2011	7	10.3	6	8.6	1	5.6
Dec 2011	2	2.9	2	2.9	1	5.6
Jan 2012	9	13.2	9	12.9	1	5.6
Feb 2012	6	8.8	6	8.6	2	11.1
Mar 2012	11	16.2	11	15.7	4	22.2
Apr 2012	3	4.4	3	4.3	1	5.6
May 2012	5	7.4	6	8.6	1	5.6
Jun 2012	7	10.3	7	10	2	11.1
Jul 2012	6	8.8	6	8.6	2	11.1
Aug 2012	2	2.9	2	2.9	1	5.6
Fisher's exact p=1.000						

C

Restricted to radiologically-confirmed pneumonia in cases

Age group at enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
1-5 months	17	25	17	24.3	4	33.3
6-12 months	25	36.8	26	37.1	3	25
13-24 months	16	23.5	16	22.9	1	8.3
25-59 months	10	14.7	11	15.7	4	33.3
Fisher's exact p=0.704						

Restricted to radiologically-confirmed pneumonia in cases

Month of enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
Aug 2011	3	4.4	4	5.7	1	8.3
Sep 2011	2	2.9	2	2.9	0	0
Oct 2011	5	7.4	6	8.6	0	0
Nov 2011	7	10.3	6	8.6	3	25
Dec 2011	2	2.9	2	2.9	0	0
Jan 2012	9	13.2	9	12.9	0	0
Feb 2012	6	8.8	6	8.6	1	8.3
Mar 2012	11	16.2	11	15.7	4	33.3
Apr 2012	3	4.4	3	4.3	1	8.3
May 2012	5	7.4	6	8.6	1	8.3
Jun 2012	7	10.3	7	10	0	0
Jul 2012	6	8.8	6	8.6	1	8.3
Aug 2012	2	2.9	2	2.9	0	0
Fisher's exact p=0.998						

D

No antibiotics by bioassay or given in hospital

Age group at enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
1-5 months	15	25.0	13	20.6	8	19.5
6-12 months	20	33.3	24	38.1	18	43.9
13-24 months	15	25.0	15	23.8	9	22.0
25-59 months	10	16.7	11	17.5	6	14.6
Fisher's exact p=0.973						

No antibiotics by bioassay or given in hospital

Month of enrolment	n Healthy controls	% Healthy controls	n URTI controls	% URTI controls	n Cases	% Cases
Aug 2011	2	3.3	4	6.3	1	2.4
Sep 2011	1	1.7	1	1.6	1	2.4
Oct 2011	2	3.3	6	9.5	2	4.9
Nov 2011	7	11.7	5	7.9	3	7.3
Dec 2011	2	3.3	2	3.2	2	4.9
Jan 2012	9	15	8	12.7	5	12.2
Feb 2012	6	10	6	9.5	4	9.8
Mar 2012	8	13.3	9	14.3	6	14.6
Apr 2012	3	5	3	4.8	2	4.9
May 2012	5	8.3	5	7.9	6	14.6
Jun 2012	7	11.7	7	11.1	3	7.3
Jul 2012	6	10	5	7.9	5	12.2
Aug 2012	2	3.3	2	3.2	1	2.4
Fisher's exact p=1.000						

On multivariable logistic regression, seen in table 8.3, not very much variation in nasopharyngeal pneumococcal concentration was associated with the variables examined, in keeping with the findings from the stratified or restricted analyses. Only the 6-12 month age group in the presence of a respiratory virus, had a significantly different (higher) pneumococcal concentration than the youngest age group, and July 2012 had samples with higher concentration than other months.

"Runny nose" also known as coryza, was part of the definition of URTI symptoms but could also be present among cases. It was present among 60/70 (86%) controls with URTI symptoms and among 33/69 (48%) pneumonia cases, $p<0.001$.

Table 8.3: Multivariable logistic regression model of differences in log_{10} nasopharyngeal pneumococcal concentration.

Univariate analysis	co-efficient	95% CI		p-value		Multivariable analysis	co-efficient	95% CI		p-value
Case-control status					*					
Healthy control	(reference)									
Control with URTI	0.43	-0.13	0.99	0.131						
Case of pneumonia	-0.38	-0.94	0.17	0.177						
Radiologically-confirmed pneumonia	0.09	-0.96	1.14	0.869						
Age group					*	Without respiratory virus				
1-5 months	(reference)					1-5 months	(reference)			
6-12 months	0.33	-0.25	0.92	0.260		6-12 months	-0.23	-1.52	1.06	0.723
13-24 months	-0.06	-0.71	0.59	0.851		13-24 months	0.40	-0.96	1.77	0.566
25-59 months	-0.05	-0.77	0.68	0.900		25-59 months	-0.82	-1.97	0.34	0.165
Presence of a respiratory virus	0.49	-0.04	1.03	0.072	*	With respiratory virus				
						1-5 months	(reference)			
						6-12 months	**0.69**	**0.01**	**1.38**	**0.048**
						13-24 months	-0.05	-0.83	0.72	0.897
						25-59 months	0.78	-0.22	1.78	0.128
Culture of pneumococcus from induced sputum	0.4834263	-0.481	1.45	0.320						
Month of enrolment					*					
August 2011	(reference)									
September 2011	0.34	-1.27	1.94	0.678			0.35	-1.31	2.01	0.679
October 2011	-0.03	-1.27	1.21	0.962			-0.03	-1.31	1.26	0.967
November 2011	0.05	-1.14	1.23	0.940			0.22	-1.02	1.45	0.731
December 2011	0.95	-0.66	2.55	0.246			0.70	-0.95	2.35	0.405
January 2012	0.48	-0.65	1.61	0.406			0.49	-0.69	1.67	0.414
February 2012	-0.41	-1.62	0.80	0.506			-0.35	-1.60	0.90	0.578
March 2012	0.11	-0.99	1.21	0.841			0.19	-0.95	1.33	0.745
April 2012	0.76	-0.66	2.18	0.294			0.85	-0.63	2.33	0.258
May 2012	0.55	-0.68	1.77	0.378			0.71	-0.55	1.97	0.268
June 2012	0.19	-0.98	1.37	0.744			-0.02	-1.23	1.20	0.977
July 2012	1.62	0.41	2.83	0.009			**1.52**	**0.25**	**2.78**	**0.019**
August 2012	0.59	-1.01	2.20	0.469			0.67	-0.96	2.29	0.418
Season										
Jan-Feb-Mar (Hot dry)	(reference)									
Apr-May-Jun (Cool wet)	0.31	-0.28	0.90	0.299						
Jul-Aug-Sept (Cool dry)	0.73	0.11	1.35	0.021						
Oct-Nov-Dec (Warm wet)	0.03	-0.59	0.64	0.930						
Prior antibiotic use	-0.60	-1.26	0.06	0.076	*					

*variables in the backwards stepwise selection process; p<0.1 in univariate analysis, plus month of enrolment and age group because cases and controls were matched for these, plus case-control status

8.4 Discussion

Pneumococcal concentration in the nasopharynx of children admitted to hospital with WHO-defined severe or very severe pneumonia was not higher than community controls with or without symptoms of URTI. In fact, control children with URTI symptoms had slightly higher nasopharyngeal pneumococcal concentration than those hospitalised with pneumonia. The difference between groups, while statistically significant, is not great, with median pneumococcal loads all within half a log of one another and considerable overlap in distributions.

This finding differs from that of Vu *et al* in Vietnam where swabs were all collected before commencing antibiotic treatment,[99] and it was initially thought this might be due to antibiotic use among children with pneumonia. Restriction of the pneumonia group to cases who had had a swab collected before commencing antibiotic therapy in hospital and who had no evidence of prior antibiotic use in a serum antibiotic bioassay, and restriction of the control groups to those who had no evidence of antibiotic use in the bioassay, only somewhat attenuated the variation between groups, with loss of statistical significance. There may be residual confounding by antibiotic use that we were not able to detect. We have previously shown that the antibiotic bioassay is more sensitive from urine than from serum, but is harder to perform, because of the difficulty in collecting urine from hospitalised sick children prior to administration of antibiotic therapy (unpublished data). PCR-based methods such as the *lytA* real-time qPCR ought to be less susceptible to alteration by antibiotic use than culture-based methods, but this may depend on timing of antibiotic use. A course of oral antibiotics in the community in the days prior to study enrolment and sample collection would be expected to have more of

an effect on qPCR estimation of nasopharyngeal pneumococcal concentration than would an intravenous antibiotic dose in hospital within minutes to hours of sample collection.

Another possible explanation of our findings is that significantly more control children with URTI symptoms than those in the pneumonia group had documented coryza, and coryza may have made it easier to detect pneumococci from a nasopharyngeal swab, so that estimation of nasopharyngeal pneumococcal concentration was relatively elevated. However, we did try to adjust for this possibility with the *alu* qPCR and, this does not explain the fact that healthy controls, none of whom had coryza, had a similar median nasopharyngeal pneumococcal concentration to the pneumonia group.

Stratifying the samples by presence of a respiratory virus, *M. pneumoniae, C. pneumoniae or B. pertussis*, did not markedly alter the distribution of the nasopharyngeal pneumococcal concentrations by group. We can suppose that a respiratory virus in the context of an URTI may have been both causing coryza and stimulating an increase in the carried nasopharyngeal pneumococcal concentration, making it easier to detect. A respiratory virus or 'atypical' bacterial pathogen in the context of pneumonia may have been causing pneumonia alone, without any pneumococcal aetiology. In the Vietnamese study presence of a respiratory virus increased pneumococcal nasopharyngeal carriage concentrations among cases of pneumonia carrying pneumococcus 15-fold, but presence of respiratory viruses among control children was not reported.[99]

Pneumococcal pneumonia may be becoming less common than is recognised, due to the introduction of the pneumococcal conjugate vaccine in Kenya in 2011 and to the amelioration of poverty in Kilifi District over time,[196] so that the WHO-defined syndrome of severe or very severe pneumonia probably did not include many cases of true pneumococcal pneumonia. In fact, this syndromic definition may have included children

with heterogenous conditions other than pneumonia but who were severely unwell, which is a limitation of this sample set.

Restricting cases to those from whom *S. pneumoniae* was grown in culture of induced sputum, or to those with radiologically-confirmed pneumonia, and so might reasonably be expected to be more likely to have had pneumococcal pneumonia, still did not reveal any statistically significant variation to the distribution of nasopharyngeal pneumococcal concentration among cases and controls. This may have been partly due to loss of power. Restricting cases to those with radiologically-confirmed pneumonia and growth of pneumococcus from induced sputum might be expected to better concentrate true cases of pneumococcal pneumonia but this would have led to further loss of power and so would need to have been done before initial sample selection. There is not much published experience with culture of sputum from young children in the diagnosis of pneumonia, but a study in Finland found that induced sputum sampling from children with pneumonia was possible, yielded mostly high quality samples, and culture of induced sputum was useful in determining likely aetiological pathogens, including *Streptococcus pneumoniae*.[154]

In a study of paediatric community-acquired pneumonia with RSV or Rhinovirus infection by Esposito et al in Italy, children with alveolar pneumonia on chest X-ray had higher nasopharyngeal pneumococcal loads by *lytA* qPCR than children with non-alveolar pneumonia, but controls without pneumonia were not studied.[150] For those with RSV infection, the mean pneumococcal nasopharyngeal concentration with alveolar consolidation was 4.5 $\log_{10}$ copies/mL compared to 3.8 $\log_{10}$ copies/mL without alveolar consolidation (p=0.04). For those with rhinovirus infection, the mean pneumococcal

concentration with alveolar consolidation was 4.2 log_{10} copies/mL and without was 3.4 log_{10} copies/mL (p=0.03).[150]

In Vietnamese children, the median pneumococcal concentration among those with radiologically-confirmed pneumonia was 7.8 x 10^6/mL, among those with other lower respiratory tract infection was 1.3 x 10^6/mL and among healthy children was 7.9 x 10^5/mL.[99] Like the study by Vu *et al* in Vietnam, the Italian study by Esposito *et al* was unable to identify a cut-off concentration to aid in the diagnosis of pneumococcal pneumonia (in this case of mixed viral-bacterial aetiology) perhaps partly because mixed viral-pneumococcal pneumonia may also present without alveolar consolidation visible on paediatric chest X-ray. The concentrations obtained by Esposito *et al* were lower than those in our study, which were in turn lower than those measured in Vietnamese children. However, it is not appropriate to directly compare concentrations obtained by real-time PCR without calibrating the PCR standards between assays so this is likely to be artefact.

In a study of HIV-infected adults with community-acquired pneumonia in South Africa, the mean nasopharyngeal log *lytA* qPCR copy number was 6.0 log_{10} copies/mL among patients with a diagnosis of pneumococcal pneumonia compared to 2.7 log_{10} copies/mL among patients without a pneumococcal diagnosis and 0.8 log_{10} copies/mL among well HIV-infected controls (p<0.001).[126] The differences between these groups are considerably greater than those in the published paediatric studies discussed above and in our study, presumably because adults, even adults with HIV infection, have a lower nasopharyngeal pneumococcal concentration than young children, in the absence of pneumococcal pneumonia.

Another South African study, in which 49% of 400 study participants with acute lower respiratory tract infection, and nasopharyngeal carriage of pneumococcus by qPCR, were children <5 years of age, 3% were older children and the remainder were adults, found that nasopharyngeal carriage density >1000 copies/mL was associated with invasive pneumococcal pneumonia.[152] The authors also found that respiratory virus co-infection was associated with nasopharyngeal pneumococcal carriage density with an adjusted odds ratio of 1.7 (95% confidence interval 1.1-2.6). This study combined adults and children, making it difficult to compare the results to other studies, and for these analyses took only subjects with lower respiratory tract infection carrying pneumococcus. The density threshold of >1000 copies/mL may have been different for children than for adults if the study had been analysed in age strata. Additionally, invasive pneumococcal pneumonia was defined by *lytA* PCR on blood, which as discussed in chapter eleven of this thesis, does not have good specificity for IPD among children.

In our study, adjusting the pneumococcal concentration in the nasopharynx for the amount of human DNA present in the swab transport medium, to allow for poor swabbing technique or copious coryza, did not substantially alter the results or the conclusions that can be drawn from them.

A striking feature of this work was the finding that almost all nasopharyngeal swabs in STGG were positive for *lytA* qPCR (98.6%) with only two swabs from cases of pneumonia and one swab from a control participant testing as negative. The corresponding nasopharyngeal/oropharyngeal swabs in UTM tested in the FTD respiratory multiplex PCR that also included a *lytA* target had a lower positivity (80.8%); this PCR also allowed positives up to 40 amplification cycles but there were differences in swab transport media, extraction methods and the PCR was multiplex. Given the widespread positivity of

nasopharyngeal swabs in STGG for *lytA* PCR, it was thought reasonable to call all swabs positive, with all swabs less than the lower limit of detection set to half that value. This supposes that all children in our population carry pneumococcus in the nasopharynx to some degree, and challenges our previous belief that carriage prevalence among children <5 years of age in Kilifi is 66%, by WHO standard culture.[168,195] We could have taken the opposite approach, and called all swabs with less pneumococcus than the lower limit of detection negative, which would have been more conservative. Of course, it is not unexpected that PCR would be more sensitive than culture for pneumococcus, since PCR can detect non-viable organisms, but it does change the paradigm to consider that all children in our population might be carriers. If carriage as defined by PCR positivity of NP-STGG, is present in all children, then it is not useful to predict transmission or invasion or progression to pneumonia. Density of carriage by qPCR is the natural progression of this chain of thought but in this study at least, nasopharyngeal pneumococcal concentration was not associated with admission to hospital with WHO-defined pneumonia among children less than five years of age. The ideal study design to answer this question would be a longitudinal study with frequent sampling, of large enough numbers to enable follow-up until pneumonia occurs; such a study is impractical.

Quantification of nasopharyngeal pneumococcal concentration does not hold promise as a useful technique for diagnosing pneumococcal pneumonia among children under the age of five years in Kenya. In our population, nasopharyngeal pneumococcal concentration among children with symptoms of upper respiratory tract infection was actually higher than among children hospitalised with pneumonia. In Vietnam and Italy, nasopharyngeal pneumococcal concentration was higher among radiologically-confirmed pneumonia cases than cases of other lower respiratory tract infection, and in Vietnam, than among healthy controls. Neither of these studies actually had a control group with

URTI symptoms for direct comparison with our study and the populations are likely to be different in other ways; Italy is a high-income country compared to Kenya and South-East Asian countries like Vietnam appear to have considerably less pneumococcal disease than countries in sub-Saharan Africa (seen in chapter eleven). All the paediatric studies mentioned, including ours, found considerable overlap in the distribution of nasopharyngeal pneumococcal concentrations by study group, in contrast to the South African studies of adults, in whom it appears to be easier to set a threshold to distinguish nasopharyngeal carriage associated with pneumococcal pneumonia.[126]

In summary, we found that among Kenyan children under five years of age, nasopharyngeal pneumococcal carriage as detected by PCR is very common and is not greater in children with WHO-defined pneumonia than among controls. Residual confounding by antibiotic use may have hampered our ability to detect high nasopharyngeal pneumococcal concentrations among pneumonia cases and the WHO syndrome of pneumonia may not have included many cases of true pneumococcal pneumonia. Carriage is likely to be augmented by the presence of upper respiratory tract viral infection, potentially accounting for the highest pneumococcal concentrations being seen among the controls with symptoms of URTI.

Chapter Nine

9 Objective two: The dynamics of pneumococcal concentration in the nasopharynx during symptomatic viral upper respiratory tract infection

9.1 Introduction

A link between viral infection and bacterial pneumonia has been established in many populations, notably during Influenza pandemics,[75,198] and ecological studies also lend support.[73] Animal models have confirmed the transmissibility and density of pneumococcal nasopharyngeal carriage increases during viral respiratory infection [82,89-91,95] and basic science experiments demonstrate the upregulation of receptors for bacterial adherence and reduced bacterial clearance during the inflammatory response associated with viral infection. [77,83-85,88,95] Influenza virus, or host factors in response to influenza virus, have been shown to disperse pneumococci from biofilms, potentiating invasion.[61]

In Peru, a recent study has found the risk of nasopharyngeal pneumococcal acquisition increased post influenza or parainfluenza acute respiratory illness in children <3 years of age.[199] While associations do not prove causality, it is widely accepted clinically that viral respiratory tract infections can lead to bacterial pneumonia. Increased nasopharyngeal pneumococcal concentration has been demonstrated in humans with pneumococcal pneumonia [126] but it is not clear whether the pneumococcal pneumonia has caused the increase in nasopharyngeal pneumococcal concentration or the other way around.

It is biologically plausible that a rise in pneumococcal nasopharyngeal concentration associated with acquisition of a viral infection could increase the likelihood of progression to pneumococcal pneumonia or invasive pneumococcal disease (IPD). If this is the case, then the concentration of pneumococci in the nasopharynx should increase following viral infection and decrease again after the viral infection is cleared. It is not clear how long the augmented risk period for development of pneumococcal disease is subsequent to viral infection. Some ecological studies have found significant associations between peak viral detection and peak IPD with a lag of one week to one month, others have found associations without applying a lag.[71-73]

We set out to study the prediction that acquisition of a respiratory viral infection is associated with an increase in pneumococcal nasopharyngeal concentration in young children, and if so, that nasopharyngeal pneumococcal concentration returns to baseline after the viral episode is over. A recent study of respiratory viral transmission among households in the rural Kilifi district of Kenya enabled us to investigate this question.[9]

Many studies are now using quantitative PCR to address differences in concentration of pathogens associated with carriage and disease states,[126,180,181] making standardization of sample collection and laboratory methodology vital.[64] This is fairly straightforward for sample types such as whole blood, where the specimen volume from which DNA is extracted is known. For samples such as nasopharyngeal secretions, the specimen volume is much more variable; all that can be known for sure is the volume of transport medium from which DNA is extracted, not the volume of the clinical specimen within the transport medium. Factors such as increased nasopharyngeal secretions (as are often seen in viral upper respiratory tract infections) may increase the volume of human specimen collected with a nasopharyngeal swab and deposited in the vial of transport

medium. Factors such as participant discomfort during swabbing, or poor training or monitoring of swab collectors may result in an inefficiently collected swab and a decreased volume of human specimen deposited in the vial of transport medium. Such differences would be expected to introduce bias, which could be corrected by an adjustment for the quantity of human DNA present in the transport medium of a nasopharyngeal swab. This is not a perfect adjustment, as it assumes a constant relationship between human DNA and specimen volume, whereas there is variation in cellularity of coryza in reality. Further data and discussion regarding the adjustment by human DNA concentration can be found in chapter five, section 5.4.

9.2 Methods

The setting for this study is a pre-existing project consisting of samples from a household study of viral transmission; "Who Acquires Infection from Whom" (WAIFW).[9] Under this parent protocol, nasopharyngeal flocked swab (NPFS) specimens were collected twice weekly irrespective of symptoms and stored in viral transport medium (VTM) during a single respiratory syncytial virus (RSV) season for six months from all members of 47 households with at least one infant and a sibling.[8] These households were rural Kenyan households in Kilifi County. A household was defined as people living in the same compound and sharing a common cooking arrangement. See figure 9.1 for results of RSV testing from one household as an example (figure supplied by Patrick Munywoki). Multiplex real-time PCR was performed on all samples to detect RSV A/B, adenovirus, rhinovirus and coronaviruses NL63, E229 & OC43 (in a sub-sample of the parent study these had been found to be the most prevalent viruses in circulation throughout the six month period)[8]. Additionally, five households were tested for influenza A/B/C, human metapneumovirus, parainfluenza viruses 1-4, and *Mycoplasma pneumoniae.* These five

households also underwent rhinovirus typing of all samples positive for rhinovirus by sequencing of the VP4-VP2 junction (unpublished data – Clayton Onyango).[200] Rhinovirus typing was necessary in order to be able to distinguish episodes of rhinovirus infection; rhinovirus was so commonly detected that without typing, strains belonging to the same episode could not be differentiated from other strains. This is demonstrated in figure 9.2.

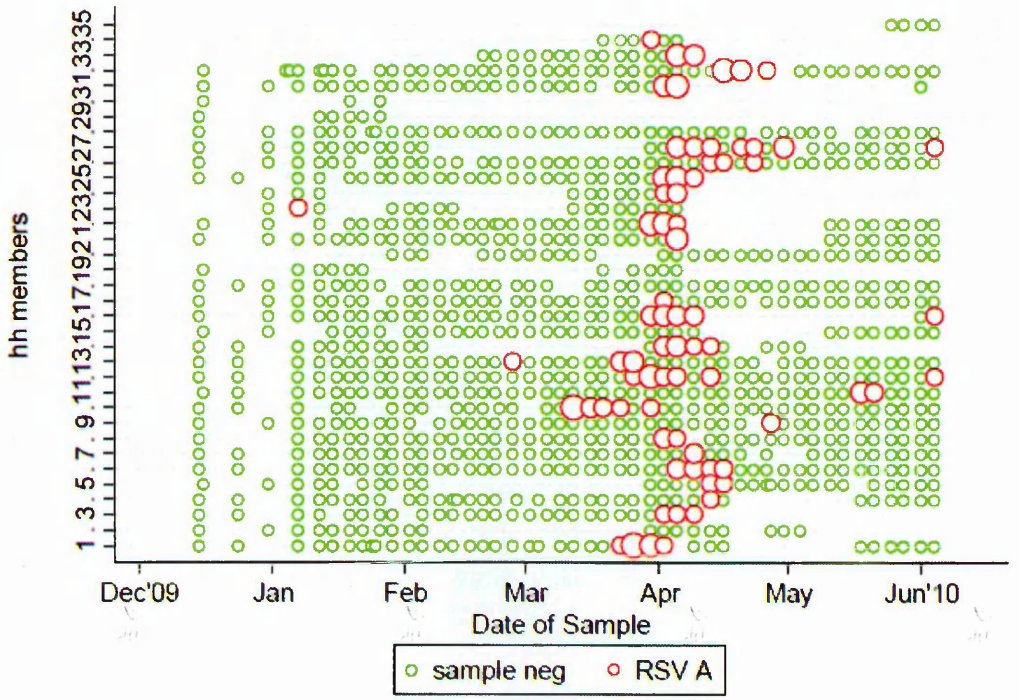


Figure 9.1: Nasopharyngeal swab sampling from all members of a rural Kenyan household twice weekly over a 6-month period. Samples negative for respiratory syncytial virus (RSV) are shown in green, samples positive for RSV are shown in red. The size of the circle corresponds approximately to the viral concentration. Figure by Patrick Munywoki.

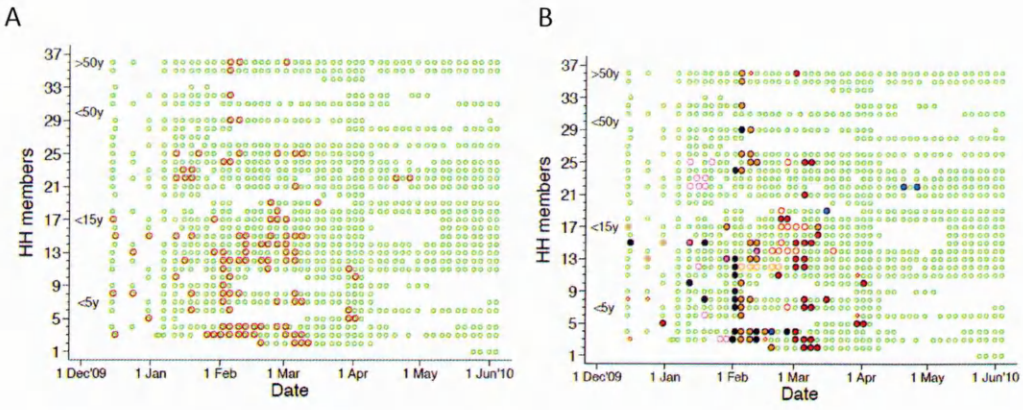


Figure 9.2: Rhinovirus typing enabled identification of episodes of rhinovirus infection for study. A: Nasopharyngeal swab sampling from all members of a rural Kenyan household twice weekly over a 6-month period. Samples negative for rhinovirus are shown in green, samples positive for undifferentiated rhinovirus are shown in red. **B:** The same household as in A; rhinovirus types are each allocated a different colour. Rhinovirus typing has allowed identification of different episodes of rhinovirus infection. Figure by Patrick Munywoki.

We selected episodes of RSV and rhinovirus infections to study. Influenza virus was not included because the sampling timeframe did not occur during the Influenza season and there were only two detected influenza episodes meeting the selection criteria. Both RSV and rhinovirus have been shown to increase pneumococcal adhesion to human respiratory epithelial cells, and while RSV has the most evidence for association with pneumonia,[5] rhinovirus is the most common upper respiratory tract virus found in most populations.

9.2.1 Defining Episodes of Viral Infection

Episodes were defined by the temporal association of symptoms of upper respiratory tract infection (URTI) with the detection of RSV or rhinovirus in at least two swabs over a period of 14 days among children under the age of five years. An episode began with the first positive swab and was terminated at the last positive swab if there were no further positive swabs from the child over 14 days. All swabs in between the positives were counted as part of the episode, regardless of positivity. To optimise specificity swabs had been defined as positive if the real-time PCR cycle threshold value for respiratory viruses was ≤35.[201]

Only episodes with symptoms of URTI at some point during the episode were considered; such symptoms were coryza, cough or sore throat. Within a single household, one child could have two or more episodes of infection with the same virus, and different children could be infected sequentially throughout the household. In this study, for each virus, only the episode from the first child to be infected per household was selected, although if more than one child became infected with the same virus simultaneously, or if a single child was the first to be infected in their household for both RSV and rhinovirus, or for more than one rhinovirus type, all of those episodes were selected. These criteria were applied to ensure that true episodes of viral infection were selected, recurrent episodes were avoided, and that pneumococcal transmission was less likely to have already occurred prior to acquisition of the viral infection, assuming pneumococcus may have been transferred along with the virus from child to child in the household. [202]

Swabs from two weeks prior to the onset of the episodes, all swabs during the episodes and two weeks after the end of the episodes were initially chosen in order to study the change in pneumococcal concentration before, during and after viral infection, but the

'after' swabs were extended to four weeks when initial pilot data was considered. A flowchart outlining the swab selection criteria can be seen in figure 3. The STATA code used to perform viral episode selection can be found in Appendix A; this code was written in conjunction with Patrick Munywoki.

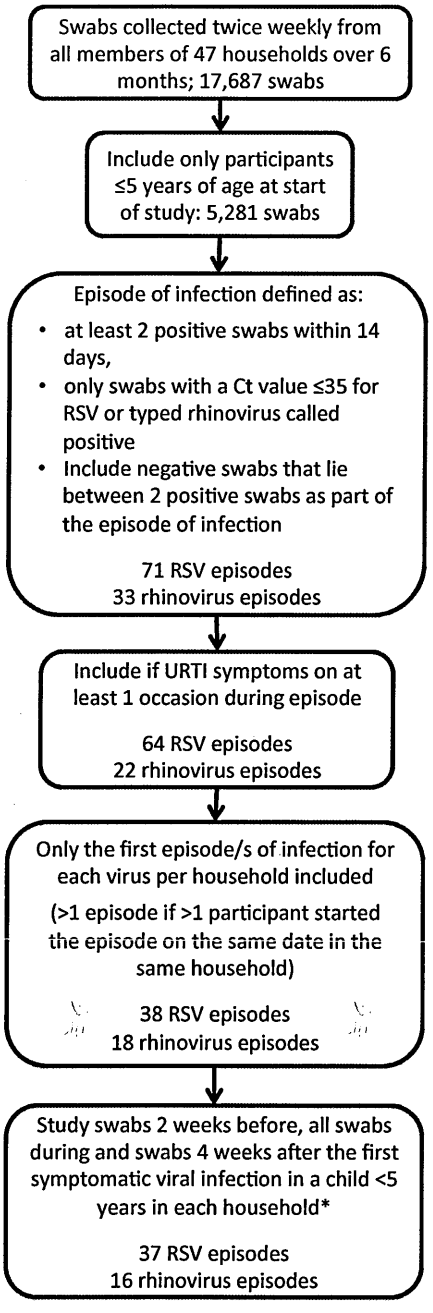


Figure 9.3: Flowchart describing the selection of viral episodes to examine from the parent study. URTI = upper respiratory tract infection. *One RSV episode was excluded due to it being a prolonged relapsing infection that by the criteria above would have meant swabs "after" the episode would still have been positive for RSV. Two rhinovirus episodes were excluded because the very first swabs collected in the study were positive so it was impossible to accurately define the episodes.

9.2.2 Laboratory Methods

Total nucleic acids had previously been extracted from all samples using the MagNA Pure LC (Roche, Indianapolis, USA) automated instrument with the Roche high performance Total Nucleic Acids kit.[8] These extracts were used to run the multiplex respiratory PCR in the parent study and stored frozen at -80°C. The same extracts were used to run the *lytA* qPCR for pneumococcal quantification and the *alu* qPCR for human DNA quantification as part of this study. These PCR methods have been described in sections 7.10 and 7.12 in chapter seven.

Pneumococcal concentration in the nasopharynx was taken as the pneumococcal concentration per µg of human DNA present in the transport medium of the nasopharyngeal swab. For the analysis, swabs with <1000 copies/mL of *lytA* detected, including swabs with no *lytA* detected, were assigned a value of 500 copies/mL before adjustment for *alu* concentration on the basis that all such swabs were below the lower limit of detection and lower limit of linearity of the pneumococcal assay.

9.2.3 Statistical methods

A pilot study of nine episodes of viral infection was carried out to inform the main analysis. Swabs during, two weeks before and two weeks after a convenience sample of episodes of symptomatic viral infection among children <5 years old were included, and both *lytA* qPCR and *alu* qPCR were carried out as above. The pilot results were summarized as a box and whisker plot of the pneumococcal concentration per µg of human DNA present in the swab transport medium and can be seen in figure 9.4. Median pneumococcal concentration before, during and after viral infection was compared using a Kruskal-Wallis rank test. Pneumococcal concentration was greater during or after viral

infection than before viral infection, p=0.022. Although there was no significant difference in pneumococcal concentration between the swabs collected during and after viral infection, viral co-infection was common and there appeared to be heterogeneity in pneumococcal concentration between individuals. It was therefore decided to extend the period for sampling after the viral infection from 2 weeks to 4 weeks for the main analysis, to concentrate on differences at the individual level, and to allow for confounding by presence of viral co-infection. In order to account for changes in nasopharyngeal pneumococcal concentration at the individual, rather than the population level, the analysis was refined. The mean natural log concentration of the "before" swabs was subtracted from the "during" swabs, for each individual viral episode. Similarly, the mean natural log concentration of the "during" swabs was subtracted from the "after" swabs for each episode. These differences, when exponentiated, yield the geometric mean fold change in pneumococcal concentration between each state. A one-sample, two-sided t-test (ie. compares the difference to nil) was used to assess statistical significance in the mean fold change, with a p-value of <0.05 considered good evidence to reject the null hypothesis. As there were only nine episodes in the pilot study, the results of this analysis were not statistically significant, but the direction of change was consistent with the hypothesis; an 11.2-fold increase in pneumococcal concentration during viral infection, p=0.0996, and a 0.5-fold decrease in pneumococcal concentration after viral infection, p=0.3246.

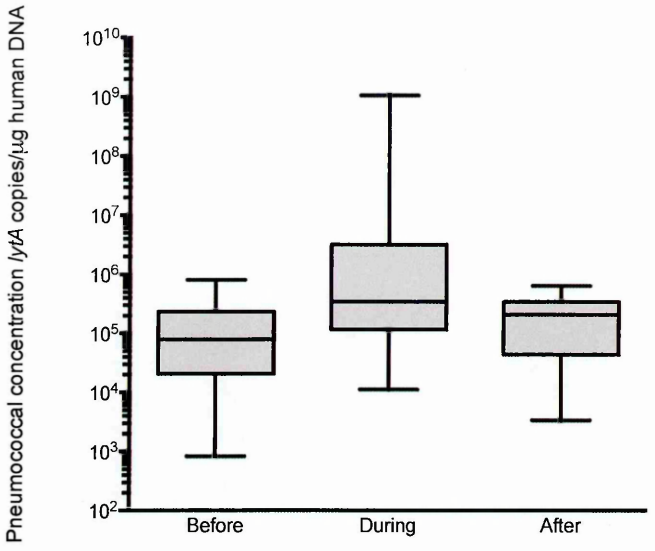


Figure 9.4: Nasopharyngeal pneumococcal concentration in *lytA* copies/µg of human DNA, before, during and after symptomatic viral upper respiratory tract infection (URTI) among nine children in a pilot study. The box and whisker plot shows the median, interquartile range and range for each swab category.

Power was calculated using a one-sample comparison of mean to a hypothesized value (in this case, 0) in STATA 11.[197] The power calculation revealed that the estimated number of episodes needed to detect a 3-fold change in pneumococcal concentration between viral infection periods, using a standard deviation in natural log pneumococcal concentration of 2.9 obtained from the pilot study, with 75% power and a significance level of 0.05, was 49. The predicted sample sizes needed for different levels of power and different effect sizes can be seen in table 9.1.

Table 9.1: Sample size calculation. Predicted sample sizes needed for different levels of power and different effect sizes.

	Power				
Fold-change in geometric mean pneumococcal concentration	**0.9**	**0.85**	**0.8**	**0.75**	**0.7**
2-fold	186	159	139	123	110
3-fold	74	63	55	49	43
4-fold	46	40	35	31	27
5-fold	35	30	26	23	21
6-fold	28	24	21	19	17

Of the potential confounders considered for the main analysis, namely the number of children in the household, the number of smokers in the household, age of the child, poverty level, cooking fuel type, whether cooking was done inside or outside, hygiene (consisting of toilet type and water source), recent use of antibiotics and presence of viral co-infection, only the last could be a confounder because it varied within episodes for individual children. Although it would have been scientifically stronger to restrict the analysis to episodes without any co-infected swabs, this would have entailed dropping almost all the data because co-infection was so common. Therefore to examine the potential role of co-infection, the individual level analysis for the main study was repeated restricted to swabs without presence of viral co-infection (rather than episodes without presence of viral co-infection).

A t-test with unequal variances was used to assess whether the differences in mean log pneumococcal concentration during and before a viral episode were significantly different for RSV and rhinovirus.

9.2.4 Ethical Clearance

The study was approved by the Kenyan National Ethical Review Committee (SSC 1932). In the original viral transmission study consent was obtained from guardians of all children studied.

9.3 Results

The main analysis consisted of 47 households (five of these households had subtyped rhinovirus episodes from the parent study), of which there were 53 viral episodes for analysis, consisting of 42 individual children and 804 sampling visits. Fifty-one viral episodes had swabs available from both before and during viral infection; 37 episodes of first RSV infection and 14 episodes of first rhinovirus infection in a household during the study period in a child <5 years of age. Of these, 28 and 12 episodes of RSV and rhinovirus infection, respectively, remained once the analysis was restricted to swabs not co-infected with other viruses. Of 804 selected sample visits, 785 had nasopharyngeal respiratory PCR results available from the parent study. For another 62 swabs, the sample was missing or had been used up in previous analyses, leaving 723 swabs available for pneumococcal PCR. By *lytA* qPCR 675/723 swabs (93.4%) were positive for pneumococcus. Of these, 23/723 swabs had insufficient sample for *alu* qPCR, and 2 were negative for human DNA by alu qPCR, leaving 698 swabs with both *lytA* and *alu* qPCR results for the main analysis. Individual plots of pneumococcal concentration before, during and after RSV and Rhinovirus infection can be seen in figure 9.5, these are illustrative examples; plots for every episode can be found in Appendix B. These plots additionally show the samples that were co-infected with another virus; 271/785 (35%) were co-infected.

The mean fold change in pneumococcal concentration, with and without adjustment for human DNA concentration for RSV and rhinovirus are in table 9.2. There was not a statistically significant difference between the viruses, p=0.250, so the mean fold change in pneumococcal concentration is also presented combined to improve power and precision. Pneumococcal concentration rose during viral infection, 3.8-fold if adjusted for the quantity of human specimen present in the transport medium, 7.5-fold if not adjusted, p<0.001 for either method. Pneumococcal concentration fell after viral infection, but this was not statistically significant. The differences in adjusted pneumococcal concentration during and before viral infection, and after and during viral infection can be seen in figure 9.6.

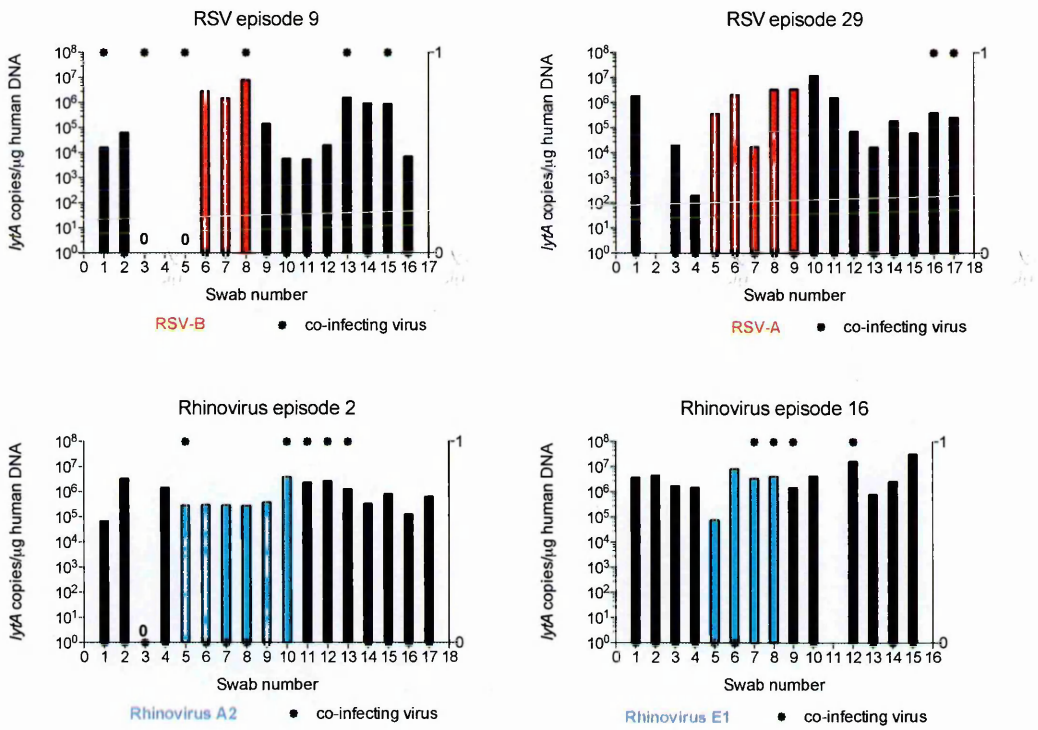


Figure 9.5: Nasopharyngeal pneumococcal concentration in *lytA* copies/µg of human DNA, before, during and after examples of episodes of symptomatic upper respiratory tract infection (URTI) with RSV or rhinovirus in children <5 years old. The y-axis of each

graph is in log_{10} scale. The swab numbers on the x-axes are consecutive twice-weekly swabs. Where a swab was not collected or the sample was unavailable, the bar is missing. Where a swab was not missing but did not have detectable pneumococcus by *lytA* qPCR, it is marked zero. For RSV episodes, swabs collected during the episode are marked in red, for rhinovirus these are blue. Swabs two weeks before and four weeks after each viral episode are black. Bars marked with an asterisk indicate swabs that were co-infected with another respiratory virus by multiplex PCR.

Table 9.2: Geometric mean fold change in nasopharyngeal pneumococcal concentration. Geometric mean fold change in nasopharyngeal pneumococcal concentration with the onset of a symptomatic upper respiratory tract infection (URTI) with RSV or rhinovirus ('during-'before' viral episode) and with the conclusion of the viral episode ('after'-'during') among children <5 years old. The analysis is shown with and without restriction to swabs without viral co-infection and with and without adjusting pneumococcal concentration for the concentration of human specimen present in the swab transport medium. There was not a statistically significant difference in mean fold change between RSV and rhinovirus episodes at the onset of the episode, so results are also presented combined.

		Differences in mean natural log *lytA* copies/µg human DNA									
		without restriction					with restriction to swabs without co-infecting viruses				
Virus	Periods studied relative to viral episode	n episodes	Geometric mean fold change	95% CI		1-sample t-test 2-sided p-value	n episodes	Geometric mean fold change	95% CI		1-sample t-test 2-sided p-value
RSV	'During'-'before' viral episode	37	**4.628**	**(1.959**	**10.935)**	**0.001**	28	**5.351**	**(1.824**	**15.693)**	**0.004**
	'After'-'during' viral episode	37	**0.477**	**(0.25**	**0.911)**	**0.026**	28	0.471	(0.161	1.378)	0.162
Rhinovirus	'During'-'before' viral episode	14	2.258	(0.865	5.89)	0.090	12	1.918	(0.490	7.500)	0.316
	'After'-'during' viral episode	15	1.332	(0.327	5.43)	0.668	12	1.260	(0.215	7.387)	0.779
Combined	'During'-'before' viral episode	51	**3.80**	**(1.95**	**7.40)**	**<0.001**	40	**3.93**	**(1.70**	**9.09)**	**0.002**
Combined	'After'-'during' viral episode	52	0.64	(0.35	1.17)	0.144	40	0.63	(0.26	1.54)	0.303
		Differences in mean natural log *lytA* copies/mL without adjustment for human DNA concentration									
Combined	'During'-'before' viral episode	51	**7.47**	**(3.28**	**16.99)**	**<0.001**	40	**9.28**	**(2.89**	**29.82)**	**<0.001**
Combined	'After'-'during' viral episode	52	0.45	(0.20	1.01)	0.053	40	**0.29**	**(0.09**	**0.90)**	**0.033**

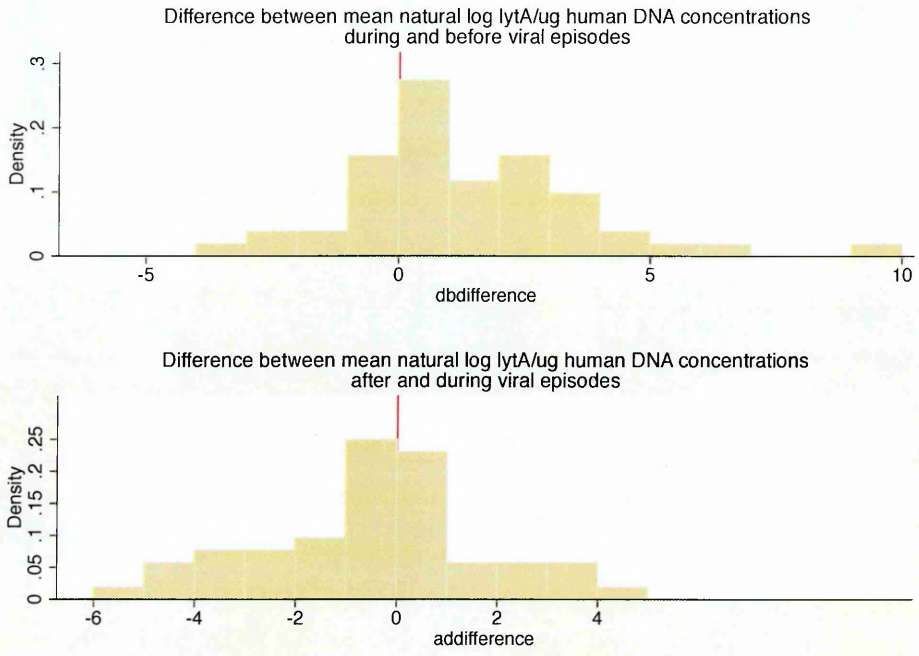


Figure 9.6: Differences in individual mean nasopharyngeal concentration. The differences between individual mean nasopharyngeal pneumococcal concentration in natural log *lytA* copies/μg human DNA 'during' and 'before' viral upper respiratory tract infection (URTI), and between concentration 'after' and 'during' viral URTI. A difference of 0 indicates no change in pneumococcal concentration with the change in viral status. Data to the right of the 0 line indicate an increase in pneumococcal concentration, data to the left indicate a decrease.

9.4 Discussion

There is considerable inter and intra-individual heterogeneity in the concentration of pneumococcus in the nasopharynx of children under the age of 5 years in rural Kenya. The main finding in this study, the 4-fold increase in nasopharyngeal pneumococcal concentration, is small but is statistically significant. If the probability of invasion

increases with an increase in pneumococcal nasopharyngeal concentration, then this 4-fold increment in bacterial concentration with viral infection may contribute to the augmented likelihood of bacterial pneumonia or invasive pneumococcal disease among children with viral infections. Additionally, the increase in bacterial concentration with viral URTI is very likely to contribute to person-to-person transmission of pneumococcal carriage.

The changes in pneumococcal concentration with viral infection appeared to be more marked for RSV than for rhinovirus, but there was not a statistically significant difference between the viruses. RSV may indeed have a stronger relationship with changes in nasopharyngeal pneumococcal concentration than rhinovirus but we were not powered to detect this.

The Peruvian study did not measure the bacterial concentration present, but did look at acquisition of pneumococcal serotypes that were not present in the nasopharynx one month prior, and found that acquisition of a new serotype of pneumococcus was associated with influenza or parainfluenza virus related acute respiratory illness but not with an RSV or rhinovirus associated illness.[199] We did not have enough episodes of influenza to look for a rise in pneumococcal concentration associated with influenza illness in our study.

Adjusting nasopharyngeal swab specimens for the concentration of human DNA present in the swab transport medium resulted in a reduction in the effect size measured. This is biologically plausible, because symptomatic viral infection can be expected to increase nasopharyngeal secretions, resulting in a greater volume of sample collected into the swab transport medium. Without adjusting for this increased volume, a falsely elevated nasopharyngeal pneumococcal concentration could be inferred from the sample. Two of

our nasopharyngeal swab samples were found to contain no human DNA; pneumococcal DNA was not detected in these samples either. These swabs were omitted from the analysis as "missing" rather than included as samples negative for pneumococcus, because it is unlikely the swabs were truly collected from the human nasopharynx, despite training and monitoring of swab collectors.

Strengths of this study include the extensive swab collection; 17,687 swabs were collected from 47 households over a six-month period, requiring considerable community co-operation. This is the first study we are aware of where rhinovirus sub-typing was available, enabling the study of rhinovirus episodes. The study would have been impossible without the sub-typing because so many different rhinoviruses were present that defining the individual episodes could not have been done otherwise. Adjusting the pneumococcal concentration for the amount of human DNA present in the sample meant that children who had a higher pneumococcal yield due to the presence of copious nasal secretions were accounted for. It also meant that poor swab technique, if problematic for some samples, would not bias the results.

This study does have some limitations. Despite high compliance from the nearly 500 occupants of the studied households, there were only 53 first viral episodes from children <5 years old available for analysis; 37 RSV episodes and 14 rhinovirus episodes had swabs collected both before and during episodes. Numbers were reduced further when the analysis was restricted to swabs without viral co-infections, as such co-infections were very common. Naturally, we could not test the swabs for all possible viral co-infections and therefore there may well have been further confounding that we could not detect.

Data on recent antibiotic use was limited, with only antibiotics given out by the study field team or local healthcare clinic recorded. However, the study field team noted that

antibiotics were generally otherwise inaccessible to the study population. It remains possible that we were unable to adequately account for this potential confounder. If this was the case, you would expect it to have an effect of potentially reducing pneumococcal concentration during the symptomatic viral episode, which would reduce our ability to see an association between viral infection and increased pneumococcal concentration.

It was beyond the scope of this study to serotype the *Streptococcus pneumoniae* isolates carried among these children, and would in any case have resulted in too much loss of power to stratify by serotype, but it is possible that the rise in pneumococcal concentration associated with viral infection could be serotype-dependent.

The effect of a modest rise in pneumococcal concentration with acquisition of symptomatic viral infection in terms of risk of transmission or invasion depends on whether the relationship between nasopharyngeal pneumococcal concentration and transmission or invasion is linear, sigmoid or more complicated. If the relationship is linear, then a 4 or 7-fold rise cannot have a great effect. If the relationship is sigmoid, there may be a threshold effect, in which case it depends where the threshold for transmission or invasion lies relative to the distribution of nasopharyngeal pneumococcal concentration in the population. If the respiratory viral infection triggers the pneumococcus to switch from opaque phase in a biofilm to transparent phase, expressing genes enabling invasion, as well as increasing the pneumococcal concentration as discussed in chapter two, this would be likely to potentiate the effect on invasive risk. How much of this increased risk of invasion is due to an increase in pneumococcal concentration and how much is due to changes in gene expression we do not know. Any increase in risk of transmission is more likely to be due to the increased nasopharyngeal pneumococcal concentration itself. We also do not fully understand the relative

contributions of serotype and strain. Certainly some serotypes are more invasive than others, and virulence factors other than capsule also play a role.[25,63]

In conclusion, there is a detectable rise in pneumococcal concentration in the nasopharynx of children who acquire symptomatic viral infection, particularly with RSV. There may be a less marked decrease in pneumococcal concentration after the episode of viral infection but we have been unable to demonstrate this conclusively. The modest increase in pneumococcal concentration may be an important contributing factor to the development of bacterial pneumonia or invasive pneumococcal disease among children with preceding viral respiratory tract infections at a population level.

Chapter Ten

10 Objective three: The dynamics of pneumococcal concentration in the nasopharynx after immunisation with pneumococcal conjugate vaccine

10.1 Introduction

Invasive pneumococcal disease (IPD) has been largely controlled in developed countries by the introduction of pneumococcal conjugate vaccines,[18,20] and encouraging data is emerging from resource-poor countries where such vaccines have been instituted. [3,12,203] The effect of the vaccine is due not only to a direct effect among vaccinated infants, but also to an indirect, or herd effect, among older children and adults who are protected by a reduction in circulating invasive serotypes; less carriage [156-159] and less transmission of vaccine-types.[160,161] Serotype replacement disease is caused by pneumococcal serotypes not covered in the vaccine in use.[20] The carriage prevalence and invasive potential of serotypes occupying the nasopharyngeal niche in place of the vaccine-types and the virulence and drug-resistance of such non-vaccine-types determine the degree of the problem that serotype replacement causes. Together with the fact that IPD is now an uncommon condition in vaccinated populations, making it more difficult to study, this means that nasopharyngeal carriage studies are an important part of monitoring the impact of pneumococcal conjugate vaccines.[164]

Detection of multiple serotype carriage can be difficult [166,170,171] and has not always been performed, hindering pre and post-vaccine surveillance efforts.[165,166] Children carry more than one serotype of pneumococcus at a time, and probably do so in up to 50% of

cases.[19,165-167] The pneumococcal microarray at St George's University Hospital in London (http://www.bugs.sgul.ac.uk/node/1) has been assessed by the PneuCarriage project (discussed in chapter two) as being one of the best methods in the world for accurate detection of multiple serotype carriage.[171]

Density of vaccine-type and non-vaccine-type carriage has seldom been reported, but semi-quantitative carriage density of vaccine-types decreased in the vaccine-arm of a PCV-7 randomised controlled trial among a Native American population [158] and among Gambian villagers in a cluster-randomised trial of PCV-7.[103] As discussed in chapter nine, we assume that nasopharyngeal colonisation density is important for both pneumococcal transmission and invasion, but we don't know whether the relationship is linear or whether there is a threshold effect. Nor do we know where the population range for carriage density lies relative to these.

In Kenya, we carried out a randomized controlled trial of the immunogenicity, reactogenicity and impact on carriage of PCV-10 among Kenyan children aged 12-59 months (the PCV-10 Reactogenicity, Immunogenicity and Safety study in Malindi [PRISM]).[10] Six months after vaccination, carriage prevalence of vaccine-type pneumococci as assessed by standard culture was 21% in recipients of two doses of PCV-10 compared to 31% in controls (p = 0.04). Previous work in Kilifi has demonstrated a seasonal increase in prevalence of pneumococcal carriage at the end of the wet season, usually around July.[195] In the PRISM study, total pneumococcal carriage prevalence by culture was 66% (95% confidence interval [CI] 59-72%) among controls at the commencement of the study in January and 72% (95% CI 66-79%) at the end of the study in July. Multiple serotype carriage and concentration of vaccine-type carriage was not assessed in the PRISM study.

Carriage prevalence is probably a marker of nasopharyngeal pneumococcal concentration, but we assume that quantification of colonisation should give a more accurate picture of the dynamics of pneumococcal carriage with vaccination.

The autolysin gene is common to all pneumococci so quantitative *lytA* PCR (qPCR) can only tell us the total concentration of pneumococcus. In this study we also asked what the concentration of carriage with vaccine-types (1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, 23F) and non-vaccine types was, and how that changed after vaccination. We aimed to answer this question using quantitative *lytA* PCR together with a pneumococcal microarray revealing the relative proportions of each serotype present. This objective was examined by studying samples from subjects enrolled in the PRISM study.[10]

10.2 Methods

In the parent PRISM study, participants were sampled by rayon nasopharyngeal swab (Medical Wire Company, UK) in skim-milk tryptone glucose glycerol (STGG) transport medium before, and at 30, 60, 90 and 180 days post PCV-10 in two different schedules. Culture was performed in real time for *S. pneumoniae, H. influenzae* and *S. aureus* in our laboratory and STGG was stored frozen at -80°C. For this sub-study, swabs collected prior to vaccination (day 0) and at days 30, 60, 90 and 180 from all children aged 12-23 months who were enrolled in dosing schedule group A (PCV-10 vaccine at day 0 and day 60) were selected; these were 47 children. One child dropped out of the study after day 30 and his swabs were excluded from the analyses.

DNA was extracted from 200μL of thawed STGG using the QIAamp DNA minikit (Qiagen, Germany) as described in section 7.6.2 in chapter seven. This nucleic acid was then used for the pneumococcal *lytA* qPCR (section 7.10 in chapter seven). Because only seven swabs were negative for *lytA* by qPCR, and three of these were positive by culture, all

swabs that tested at or below the limit of detection for *lytA* qPCR (1000 copies/mL) were set at 500 copies/mL, including the negatives. As well as the *lytA* qPCR, STGG samples underwent a quantitative *alu* PCR (section 7.12, chapter seven) for human DNA to adjust the *lytA* concentration for the proportion of transport medium that was actual human specimen. The pneumococcal PCR was then expressed in copies of *lytA* per μg of human DNA.

In addition, STGG was cultured on streptococcal selective agar (Oxoid, UK) and the resulting growth harvested and suspended in phosphate-buffered saline for further DNA extraction using the QIAamp DNA minikit with a pre-lysis step (section 7.6.3, chapter seven). This DNA was used for the pneumococcal microarray, described in section 7.13 of chapter seven. The microarray was carried out at St George's, University of London. At St George's, they have tried using the microarray without the culture step, extracting DNA directly from STGG and amplifying it before proceeding with the microarray, but have found that sensitivity for detecting minor serotypes is lost with that method, and so recommend the culture step (personal communication, Jason Hinds, St George's University Hospital).

10.2.1 Analysis

The pneumococcal microarray gives the relative abundance of each serotype present and considerably enhances detection of multiple serotype carriage; the *lytA* qPCR yields the total concentration of pneumococcus. It is then possible to calculate vaccine-type (VT) and non-vaccine-type (NVT) pneumococcal nasopharyngeal concentration for each swab for which there is microarray serotype data, using the proportion for each serotype (from the microarray) of the total concentration (by *lytA* qPCR). Swabs without growth of pneumococci for microarray typing were missing the VT and NVT proportions but the

total concentration by *lytA* qPCR was available, as PCR was more sensitive than culture. In order to better understand the missing data, the pneumococcal concentration by *lytA* qPCR was compared for samples that grew and did not grow pneumococcus, and variables potentially associated with growth of pneumococcus were examined by logistic regression with pneumococcal concentration in strata formed by quintiles.

The distribution of the total pneumococcal concentration at each time point was examined using the results of the *lytA* qPCR (adjusted for human DNA concentration) in all swabs.

In order to assess changes in VT and NVT pneumococcal concentration over time post-vaccination, a range of methods was explored in order to address limitations of the dataset. These limitations included missing VT and NVT data for samples that did not grow pneumococcus on culture for microarray but were *lytA* qPCR positive; many 0 values in VT or NVT concentration eg. from swabs that contained only NVT (and therefore "0" VT) or swabs that contained only VT (and therefore "0" NVT) and the timing of the study over January to July being in line with seasonal increases in pneumococcal prevalence so that we may have confounding by season.

1) A visual display of all data. Firstly all data was displayed visually; stacked area charts to present all serotype data for each individual study participant, and the log *lytA* concentration for all pre-vaccine and combined post-vaccine swabs.

2) Changes in VT and NVT pneumococcal concentration with time, for swabs with microarray data. Secondly, VT and NVT pneumococcal concentrations, from swabs that had microarray serotype data, were studied by time point with Wilcoxon signed rank tests used to compare median concentrations. The differences between concentrations of

pre-vaccine swabs and paired swabs collected six months later were calculated and analysed using a sign test. Total pneumococcal concentration was analysed in the same way, but from all swabs, using *lytA* qPCR data alone, as microarray serotype data was not needed.

3) Changes in VT and NVT pneumococcal concentration with time, with imputed data for swabs missing microarray data. Data were re-analysed after imputing the proportions of VT and NVT for swabs positive by *lytA* qPCR but without serotype data from microarray. To impute data, the median VT concentration over the median NVT concentration among swabs yielding VT or NVT pneumococcus at each time point was used to generate a VT/NVT ratio for each time point. This ratio was then applied to the total pneumococcal quantity by *lytA* qPCR to yield the relative VT and NVT quantities for 58 swabs that were missing serotype data. The median VT and median NVT concentrations of all swabs that underwent microarray were not used, because several of the median VT concentrations by time point were 0.

4) The ratio of median VT to median NVT concentration. Changes in VT and NVT concentrations over time were probably confounded by seasonality. To avoid this, a different approach using the ratio of VT to NVT pneumococci was utilized, strengthening the analytic findings. Taking the ratio of the median VT concentration to the median NVT concentration at each time point described in 3) above, the change in log VT:NVT ratio by time point was examined using linear regression.

5) Individual ratios of VT:NVT concentration. Individual ratios of VT:NVT concentration were determined for study participants with multiple carriage of pneumococcal serotypes that included both VT and NVT serotypes. A Wilcoxon signed-rank test was used to

compare the median ratios between day 0 and each other swab collection time point until day 180 post-vaccine.

6) Serotype-specific analyses. The serotype-specific pneumococcal concentration at each time point was examined for the six most common serotypes, a Wilcoxon signed-rank test used to compare paired median concentrations by time point and a rank-sum test used to compare median pre and post-vaccine concentrations.

The microarray, in addition to providing the relative abundance of each pneumococcal serotype present, assesses the genome backbone of each serotype. A subtype for each serotype is assigned based on genes in the genome backbone, independent of the capsular locus. Further details on the microarray analysis can be found in section 7.13 of chapter seven.

Finally, the analysis of changes in VT, NVT and total pneumococcal concentration from swabs that had microarray serotype data (analysis 2, above) was repeated without adjusting pneumococcal concentration for the concentration of human DNA present in the nasopharyngeal swab transport medium.

10.2.2 Ethical Clearance

The study was approved by the Kenyan National Ethical Review Committee (SSC 1932). In the original PRISM study consent was obtained from guardians of all children studied.

10.3 Results

Figure 10.1 details the samples available for study. By *lytA* qPCR, 97% of all swabs were positive so, for the purposes of analysis, all swabs were assumed to be positive at all time

points, at varying concentrations. By microarray, which involved an initial culture step, 36/46 (78%) subjects were carrying pneumococcus at day 0, 40/45 (89%) at day 30, 27/46 (59%) at day 60, 33/46 (72%) at day 90 and 36/45 (80%) were carrying pneumococcus at day 180.

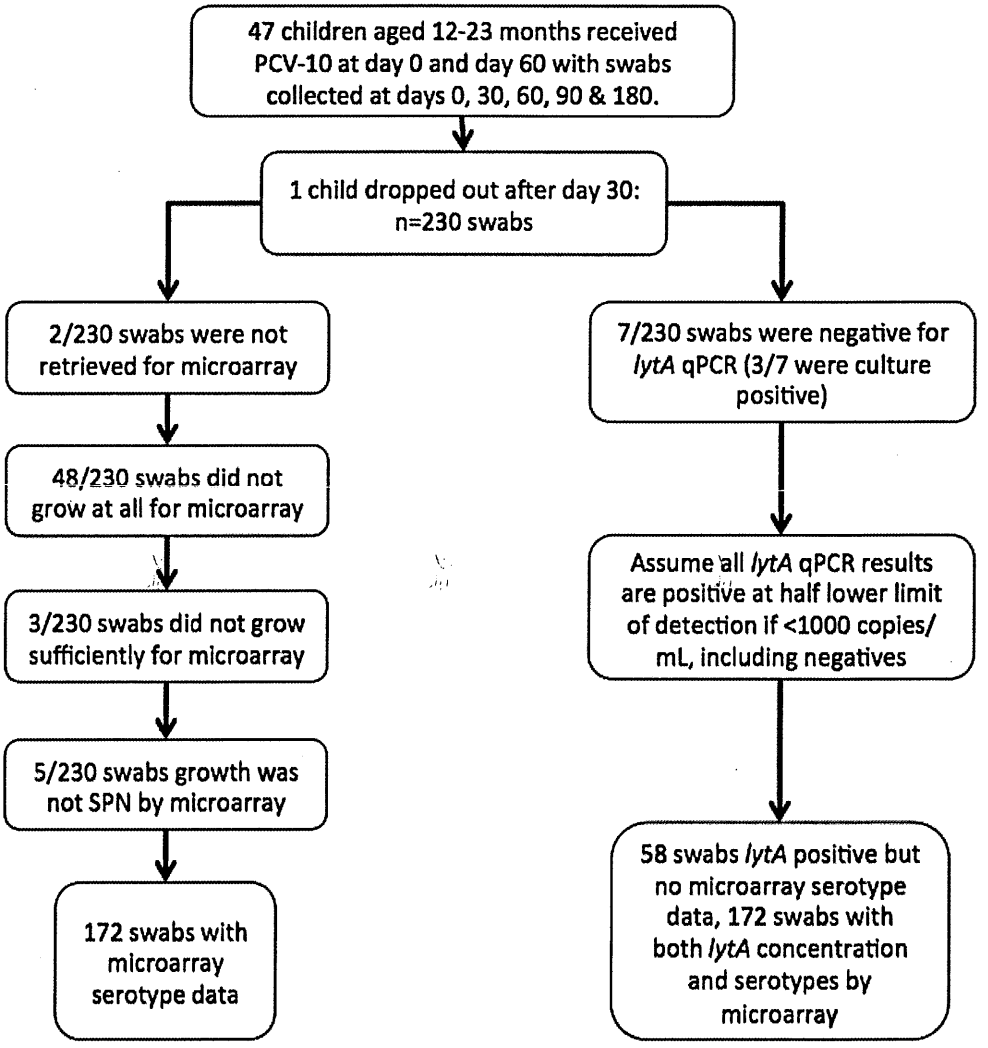


Figure 10.1: Samples available for study

The proportions at each time point carrying any vaccine-type, only non-vaccine-types, and no pneumococcus at all, by microarray, can be seen in table 10.1. As this was a sub-study of the original PRISM study three years earlier, in which nasopharyngeal swabs had been cultured for pneumococcus using the WHO standard method [168] in our laboratory, we also had those original results with which to compare the culture step for microarray. The culture for microarray and the original swab culture three years previously both yielded pneumococcus in 154/230 swabs, both were negative for pneumococcus for 47/230 swabs, and were discordant for 29/230 swabs. Of the discordant swabs 11 grew pneumococcus originally but failed to do so for microarray and 18 swabs yielded pneumococcus by culture for microarray but had not done so originally by the standard WHO method.

Table 10.1: Proportion of swabs carrying any vaccine-type, only non-vaccine-type, or no pneumococcus, by culture step and microarray analysis

Time point (days)	n	Vaccine-type pneumococcus		Non-vaccine-type pneumococcus		No pneumococcus	
0	46	18	39%	18	39%	10	22%
30	45	18	40%	22	49%	5	11%
60	46	12	26%	15	33%	19	41%
90	46	17	37%	16	35%	13	28%
180	45	17	38%	19	42%	9	20%

Of the 154 swabs that grew pneumococcus on both occasions, 139 were concordant serotypes and 15 were discordant serotypes. Of the 15 discordant serotypes, nine were concordant to serogroup level and six were not. Therefore 148/154 (96%) swabs with growth of pneumococci were concordant to at least serogroup.

The swabs that did not yield viable pneumococci for serotyping by microarray had lower pneumococcal concentration by *lytA* qPCR (median 4.04 log copies/μg human DNA, IQR 3.51-4.55 log copies/μg human DNA) than the swabs that did grow pneumococci, (median 5.19 log copies/μg human DNA, IQR 4.60-5.92 log copies/μg human DNA) p<0.0001. This is illustrated in figure 10.2.

In multivariable logistic regression; gender, age at time of swab collection, recent antibiotic use (within 30 days), month of swab collection, number of smokers in the household, mid-upper arm circumference, source of drinking water and type of cooking fuel were not associated with culture positivity but pneumococcal concentration by *lytA* qPCR was significantly associated (data not shown). The proportion culture positive in strata of *lytA* concentration can be seen in figure 10.3.

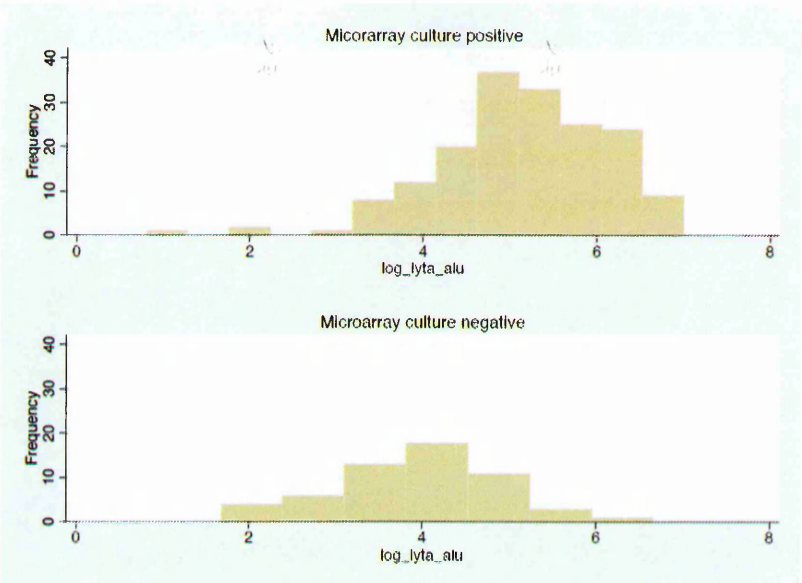


Figure 10.2: Distribution of nasopharyngeal pneumococcal concentration in log_{10} *lytA* copies/μg human DNA for swabs that did and did not grow pneumococci on culture in preparation for microarray.

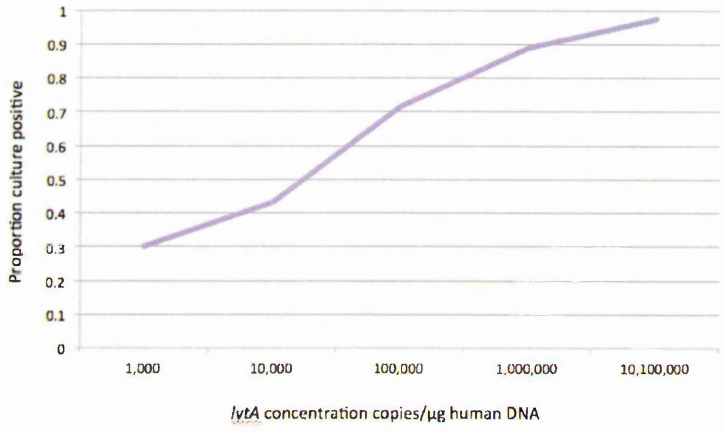


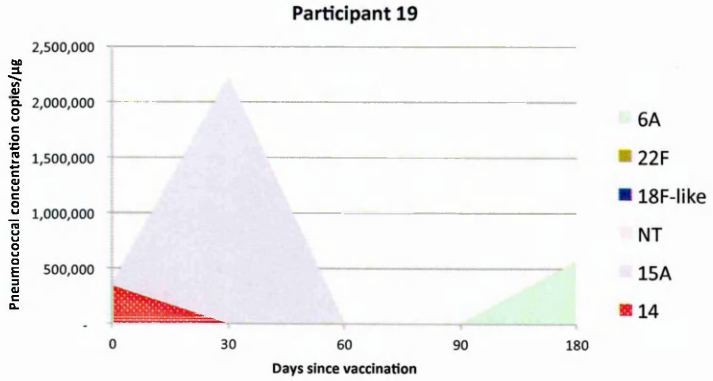
Figure 10.3: Proportion of nasopharyngeal swabs that were positive on culture for pneumococcus by five strata of pneumococcal concentration by *lytA* qPCR.

In order to assess changes in total, VT and NVT pneumococcal concentration over time post-vaccination, a range of methods was explored; here the results are presented in the same order as the analytic methods were discussed above.

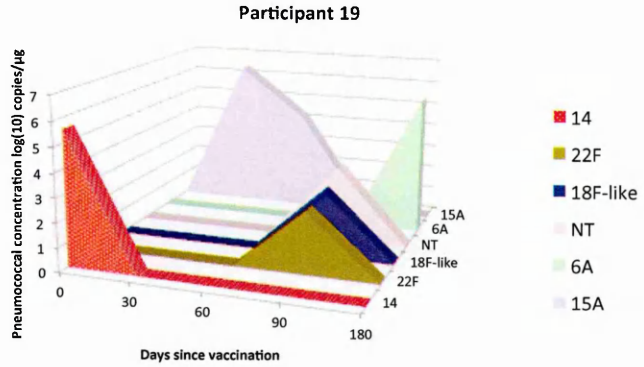
1) A visual display of all data. Examples of the visual representation of serotype-specific nasopharyngeal pneumococcal concentration across time, by individual study participant, can be seen in figure 10.4. There are two graphs, A and B, presented for each individual child. Graphs marked "A" are of nasopharyngeal pneumococcal concentration in *lytA* copies/µg of human DNA over the six months post vaccination and demonstrate the dominance of the principal serotype and the impressive changes in concentration that occurred within individuals. Graphs marked "B" are of $\log_{10}$-transformed nasopharyngeal pneumococcal concentration in $\log_{10}$ *lytA* copies/µg of human DNA and reveal the degree of multiple carriage that occurred and the relative concentrations between individuals.

"Unknown" serotypes seen on these charts are indicative of swabs for which the *lytA* qPCR was positive, but culture for microarray was not, so that we are lacking serotype data for those swabs. Graphs for every study participant can be found in Appendix C.

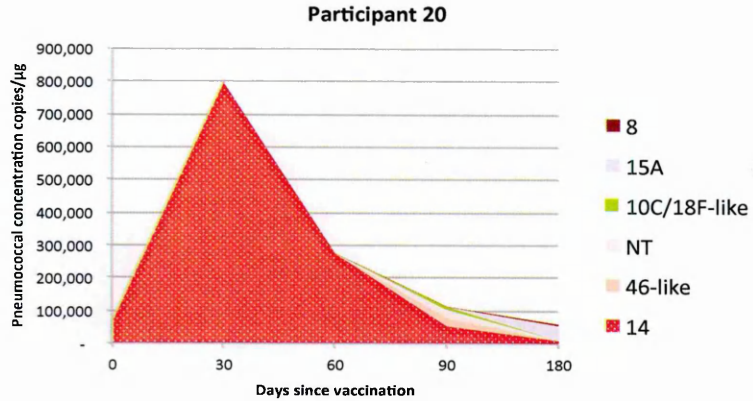
The log *lytA* concentration for all day 0, pre-vaccine swabs, and the log *lytA* concentration distribution for all day 30, 60, 90 and 180, post-vaccine swabs can be seen in figure 10.5.


Participant 19
Pneumococcal concentration copies/µg
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
6A
22F
18F-like
NT
15A
14

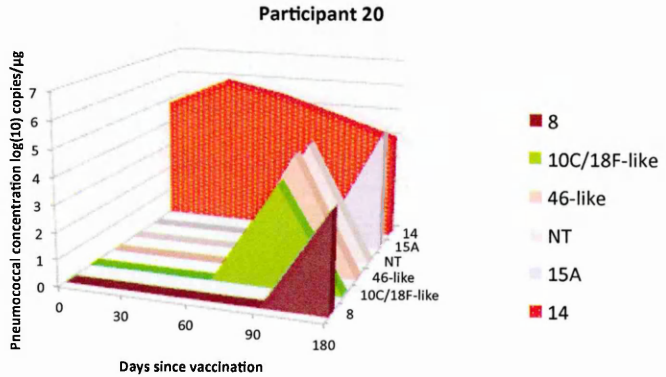
A


Participant 19
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15A
6A
NT
18F-like
22F
14
14
22F
18F-like
NT
6A
15A

B


Participant 20
Pneumococcal concentration copies/µg
900,000
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
8
15A
10C/18F-like
NT
46-like
14

A


Participant 20
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
14
15A
NT
46-like
10C/18F-like
8
8
10C/18F-like
46-like
NT
15A
14

B

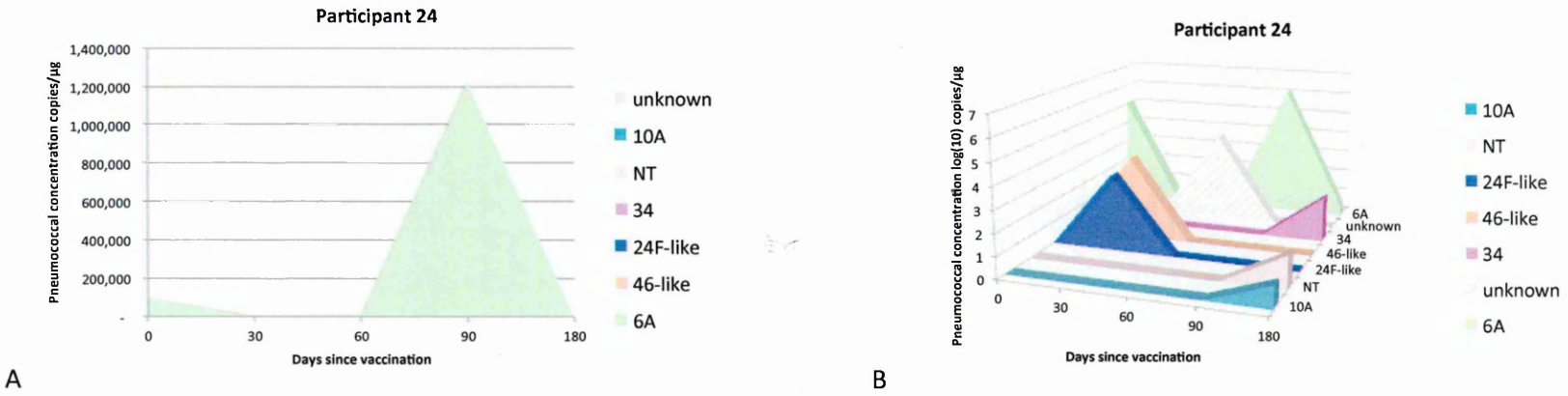


Figure 10.4: Nasopharyngeal pneumococcal concentration over six months post vaccination with 10-valent pneumococcal conjugate vaccine (PCV-10) in three individual study participants. The full set of graphs for all 46 children is available in Appendix C. Where swabs failed to grow on culture for microarray serotyping, thus serotype data is not available, the pneumococcal concentration of the unknown serotype/s is marked in grey stripes. PCV-10 vaccine-types (1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, 23F) are marked in stippled colours, non-vaccine-types are in plain colours. NT = not typable. Each study participant has two graphs; A and B. **A:** Pneumococcal concentration in *lytA* copies/µg of human DNA. Note that the y-axes are each to a different scale, due to the wide variation in pneumococcal concentration. Many of the minor serotypes cannot be seen due to the dominance of the major serotype, which can clearly be seen. **B:** Log_{10}-transformed pneumococcal concentration in log_{10} *lytA* copies/µg of human DNA. The same data as in the A graphs, log_{10}-transformed, reveals the diversity of multiple serotype carriage.

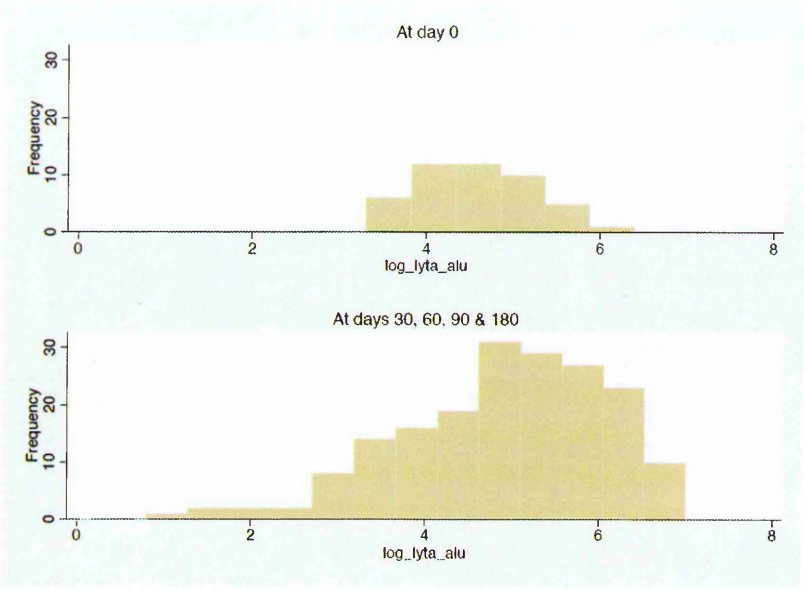
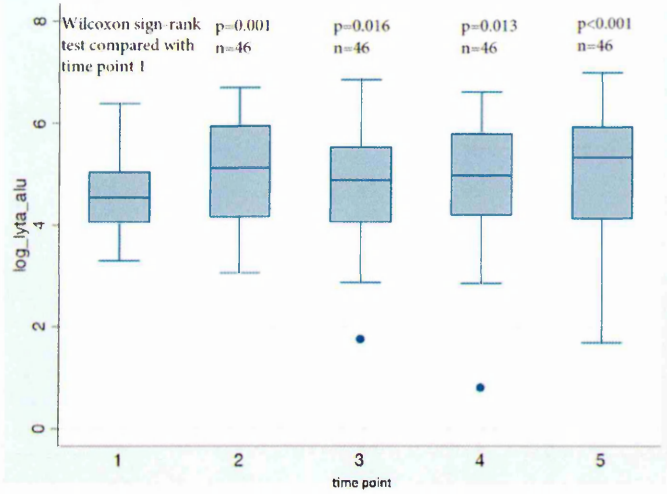


Figure 10.5: Pneumococcal concentration, in $\log_{10}$ *lytA* copies/µg human DNA, in 230 nasopharyngeal swabs from 46 children before vaccination (day 0) and after vaccination (days 30, 60, 90 and 180 since vaccination) with PCV-10.

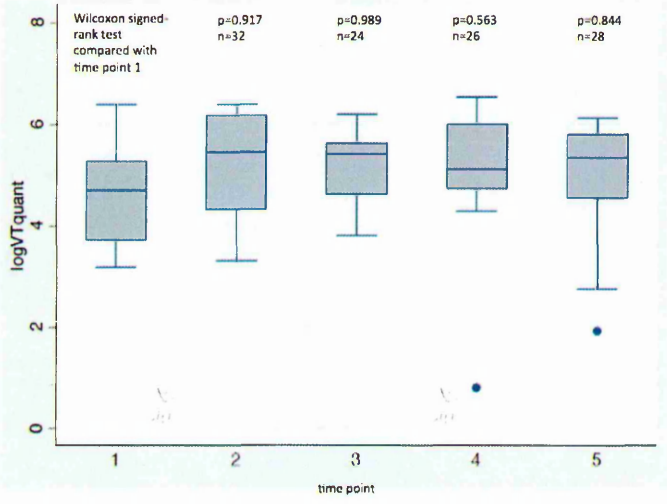
2) Changes in VT and NVT pneumococcal concentration with time for swabs with microarray data. Log-transformed VT, NVT and total pneumococcal concentrations over time can be seen in figure 10.6; data are log-transformed for ease of display but p-values are a non-parametric paired analysis of untransformed data, so as not to lose the zero quantities. Here the paired analysis of individuals at day 0 and day 30, at day 0 and day 60, at day 0 and day 90 and at day 0 and day 180 reveals no significant changes in VT pneumococcal concentration, but significant increases in NVT pneumococcal concentration. Total pneumococcal concentration likewise increases significantly over time. The analysis for VT and NVT concentration is restricted to data from swabs that were culture positive for microarray, in order to assign serotype, and does not include

data from swabs that were *lytA* qPCR positive but microarray negative. The differences in VT, NVT and total pneumococcal concentration between day 0 and day 180 are displayed in figure 10.7 and a sign test demonstrates no change in VT concentration, a significant increase in NVT concentration and an overall increase in pneumococcal concentration over the six-month period post vaccination.

A Total pneumococcus



B Vaccine-type pneumococcus



C Non-vaccine-type pneumococcus

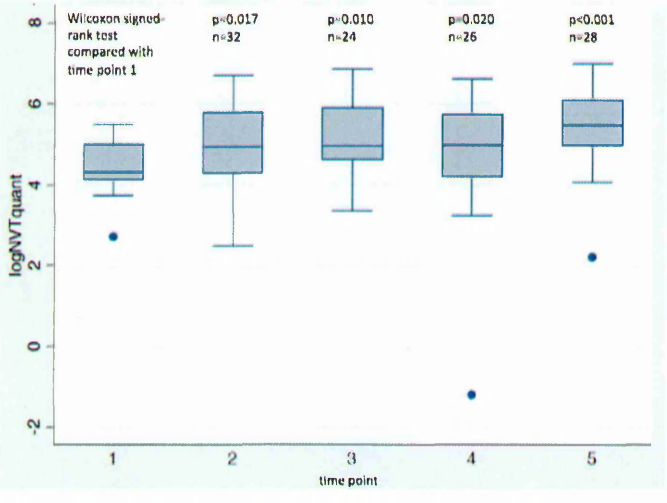


Figure 10.6: Log_{10}-transformed total, vaccine-type and non-vaccine type nasopharyngeal pneumococcal concentration over time. Time point 1 = day 0, time point 2 = day 30, time point 3 = day 60, time point 4 = day 90 and time point 5 = day 180 since vaccination with PCV-10. Data were log_{10}-transformed for display only. P-values were a non-parametric paired analysis of untransformed data. Pneumococcal concentration was in *lytA* copies/µg human DNA.

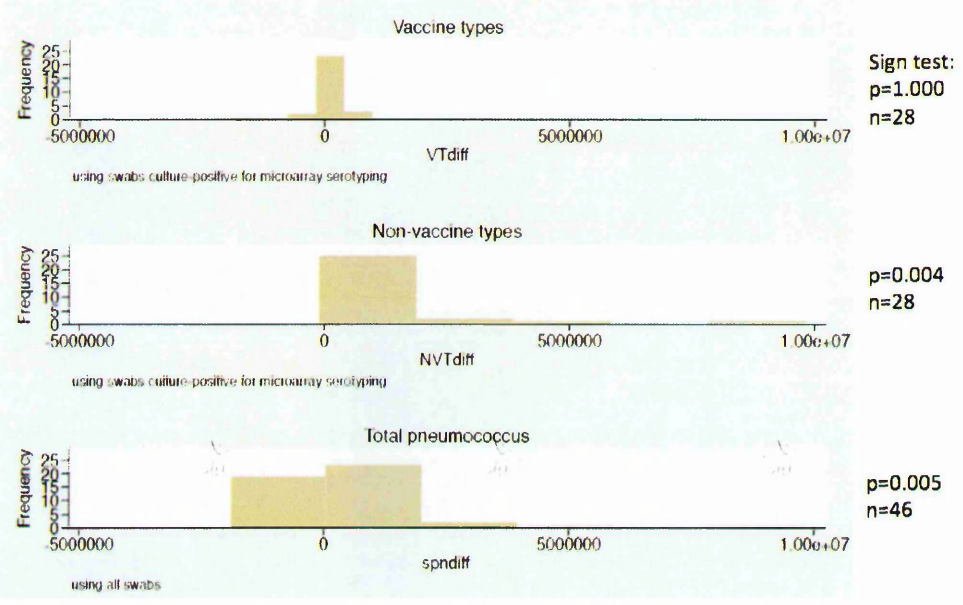


Figure 10.7: Differences in vaccine-type, non-vaccine-type and total nasopharyngeal pneumococcal concentration between day 0 and day 180 post-vaccination with PCV-10. Pneumococcal concentration was in *lytA* copies/µg human DNA.

3) Changes in VT and NVT pneumococcal concentration with time with imputed data for swabs missing microarray data. The analysis was repeated with imputed data for the 58 swabs that were positive by *lytA* qPCR but microarray serotype data was not available.

The ratio of median VT concentration over the median NVT concentration among swabs yielding VT or NVT pneumococcus at each time point was used to impute data. Several of the median VT concentrations by time point were 0, so the median VT and median NVT concentrations of all swabs that underwent microarray were not used, as can be seen in tables 10.2 and 10.3. The Wilcoxon signed-rank paired analysis of the dataset including imputed data again revealed no change in VT pneumococcal concentration and a significant increase in NVT pneumococcal concentration with time post-vaccination. Figure 10.8 shows the changes in VT and NVT concentration over time for all 46 swabs including the imputed data.

Table 10.2: Median vaccine-type (VT) and non-vaccine-type (NVT) nasopharyngeal pneumococcal concentration among swabs that had vaccine-types detected by microarray analysis and swabs that had non-vaccine-types detected by microarray analysis, which involves an initial culture step. Pneumococcal concentration was in *lytA* copies/µg human DNA.

Days since vaccination	N (VT)	median VT concentration	N (NVT)	median NVT concentration
0	18	51,372	24	21,219
30	18	284,397	31	88,520
60	12	260,145	20	94,808
90	17	129,922	25	99,523
180	17	221,241	27	293,745

Table 10.3: Median vaccine-type (VT) and non-vaccine-type (NVT) nasopharyngeal pneumococcal concentration among all swabs that underwent microarray analysis. Pneumococcal concentration was in *lytA* copies/µg human DNA

Days since vaccination	N (VT)	median VT concentration	N (NVT)	median NVT concentration
0	36	756	36	13,486
30	40	0	40	35,199
60	27	0	27	55,063
90	33	6	33	18,047
180	36	0	36	117,193

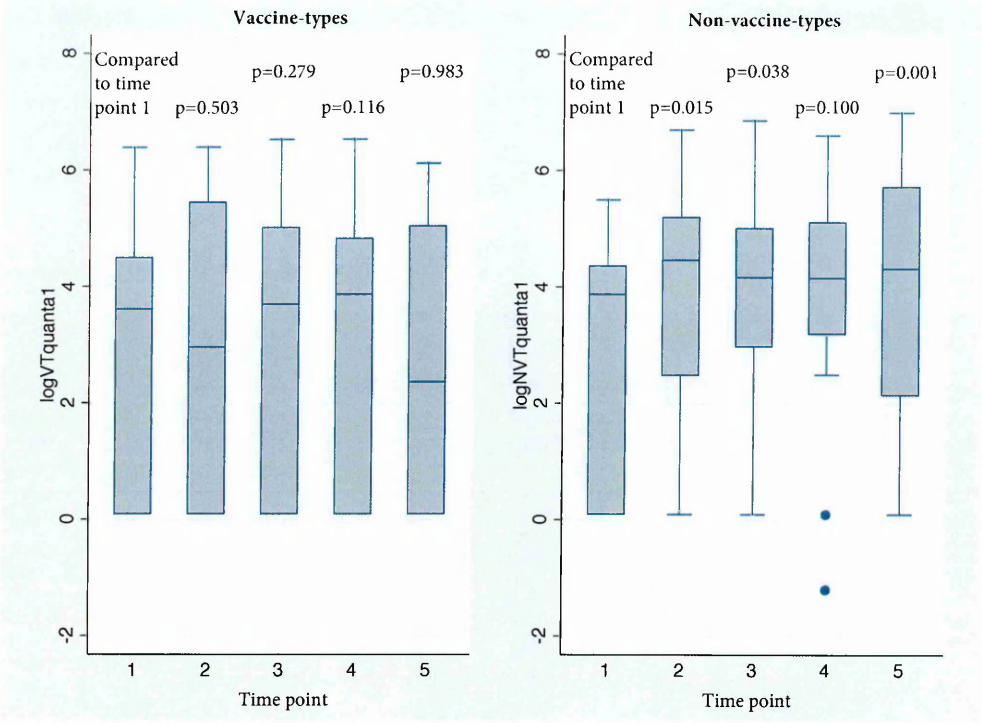


Figure 10.8: Log_{10}-transformed vaccine-type (VT) and non-vaccine-type (NVT) nasopharyngeal pneumococcal concentration over time, including imputed data.

Imputed data was for swabs with total pneumococcal concentration available by *lytA* qPCR, but proportion VT and NVT not available because of no or insufficient growth on culture for microarray analysis. Time point 1 = day 0, time point 2 = day 30, time point 3 = day 60, time point 4 = day 90 and time point 5 = day 180 since vaccination with PCV-10. Zero values were set to 0.1 and data were log_{10}-transformed for display only. P-values were a Wilcoxon signed-rank paired analysis of untransformed data. Pneumococcal concentration was in *lytA* copies/µg human DNA.

4) The ratio of median VT to median NVT concentration. An alternative method of analysis looked at the ratio of VT to NVT concentrations. This could be done overall for the whole dataset, taking a ratio of the median VT concentration to the median NVT

concentration for swabs yielding VT or swabs yielding NVT, or on an individual basis using individual participants with multiple serotype carriage. The ratio of median VT:NVT concentrations fell by 69% over the six months post vaccine and this can be seen in figure 10.9. The log of this ratio declined significantly with a p-value of 0.036.

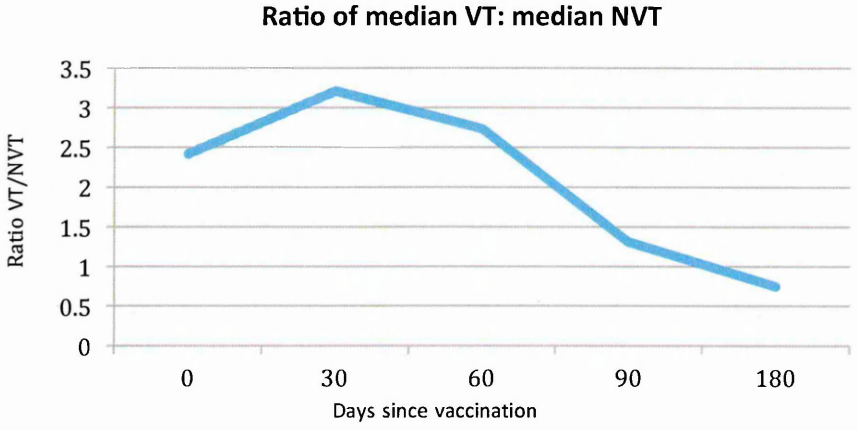


Figure 10.9: The ratio of median VT to median NVT pneumococcal concentration in the nasopharynx. The median values were taken from swabs with VT pneumococcal growth or NVT pneumococcal growth on culture for microarray analysis at each time point since PCV-10 vaccination.

5) Individual ratios of VT:NVT concentration. There were 65 swabs with multiple serotype carriage, and of those, 37 had carriage of both VT and NVT pneumococci. These 37 swabs differed from the entire dataset only in that they were more likely to come from homes containing multiple smokers and from participants with lower mid-upper arm circumference (MUAC). This can be seen in table 10.4. A box plot of the log ratio of VT to NVT quantities can be seen in figure 10.10. Statistical significance was not demonstrated.

Table 10.4: Characteristics of swabs containing both vaccine-type (VT) and non-vaccine-type (NVT) pneumococcal carriage compared to all other nasopharyngeal swabs.

	VT & NVT multiple carriage		All other swabs		
	n/37	(%)	n/193	(%)	p-value
Days since vaccination					0.717
Day 0	6	16	40	21	
Day 30	9	24	37	19	
Day 60	5	14	41	21	
Day 90	9	24	37	19	
Day 180	8	22	38	20	
Female	14	38	91	47	0.298
Firewood as cooking fuel type	37	100	193	100	NA
Source of drinking water §					0.395
Borehole	0	0	5	3	
Dam	13	35	52	27	
Pond	6	16	14	7	
Protected well	1	3	14	7	
Rainwater	0	0	5	3	
Stream/river	2	5	8	4	
Unprotected well	15	41	95	49	
Smokers in household §					0.003
none	30	81	135	70	
1	1	3	39	20	
2	3	8	17	9	
3	3	8	2	1	
Antibiotic use within 30 days	0	0	15	8	0.138
	mean	(sd)	mean	(sd)	p-value
Age (months)	20.0	4.4	20.4	3.8	0.597
MUAC (cm)¶	14.4	0.99	14.9	1.16	0.021

¶ mid-upper arm circumference, § Fisher's exact

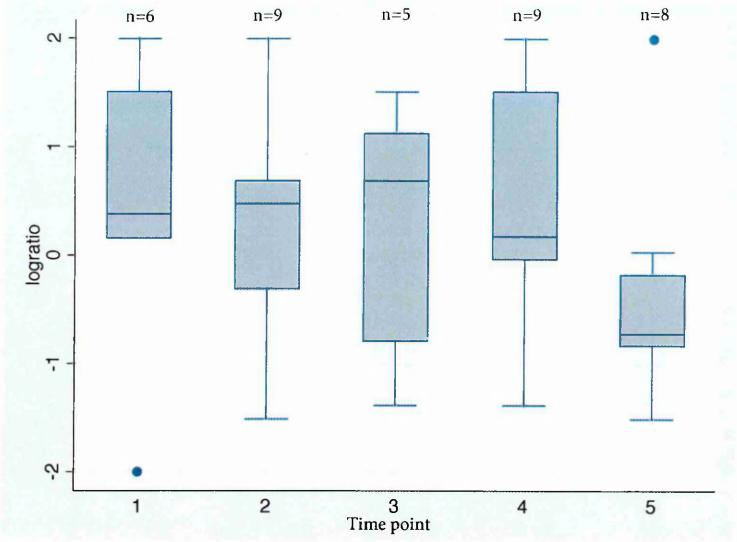


Figure 10.10: Log_{10} ratio of vaccine-type (VT) to non-vaccine-type (NVT) pneumococcal concentration in the nasopharynx among 37 swabs carrying both VT and NVT pneumococcus, by time point. Time point 1 = day 0, time point 2 = day 30, time point 3 = day 60, time point 4 = day 90 and time point 5 = day 180 since vaccination with PCV-10.

6) Serotype-specific analyses. Serotype concentration over time for the six most common serotypes, 6A, 19F, 23F, 15B, 14 and 6B, can be seen in figure 10.11. The numbers for each serotype were small and statistically significant changes were not detected for any of the individual serotypes.

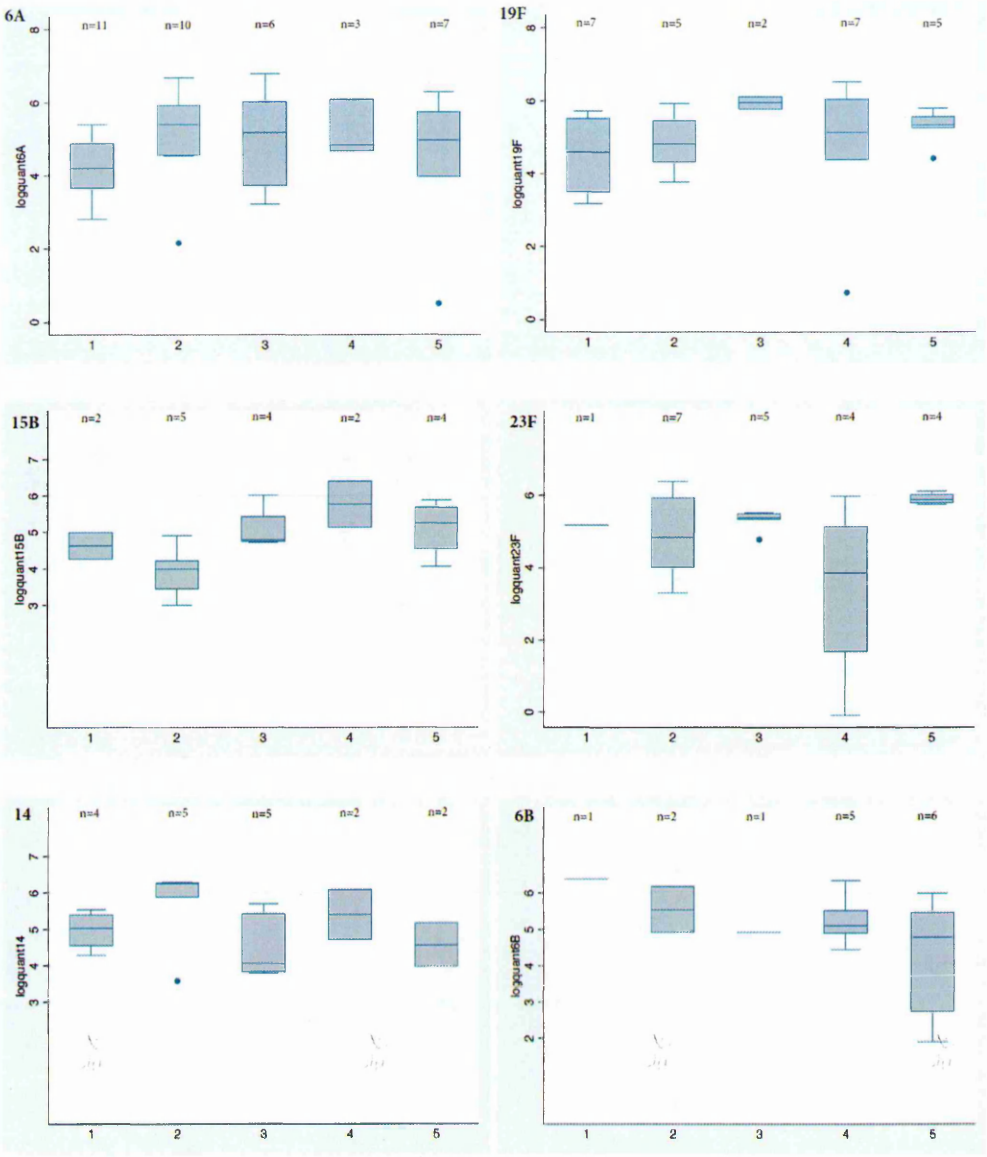
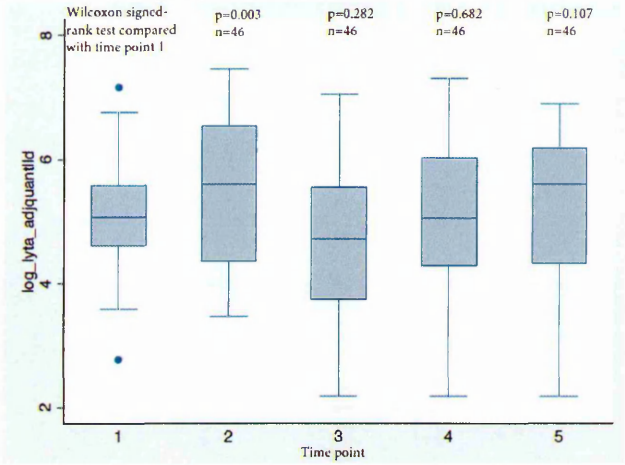


Figure 10.11: Nasopharyngeal pneumococcal log_{10} concentration by the six most common individual serotypes among over time. Time point 1 = day 0, time point 2 = day 30, time point 3 = day 60, time point 4 = day 90 and time point 5 = day 180 since vaccination with PCV-10. Pneumococcal concentration was in *lytA* copies/µg human DNA.

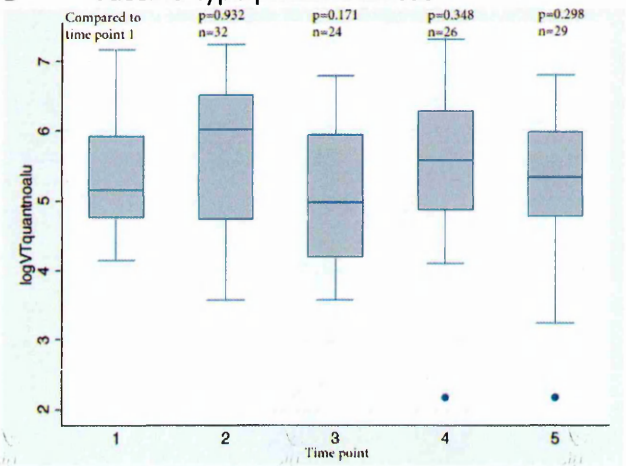
Repeating analytic method 2), without adjusting pneumococcal concentration for human DNA present in the swab transport medium, revealed no significant changes in VT pneumococcal concentration, a significant increase in NVT pneumococcal concentration,

and no significant change in total pneumococcal concentration over time post vaccination. Log-transformed VT, NVT and total pneumococcal concentrations over time, without adjustment for human DNA concentration, can be seen in figure 10.12.

A Total pneumococcus



B Vaccine-type pneumococcus



C Non-vaccine-type pneumococcus

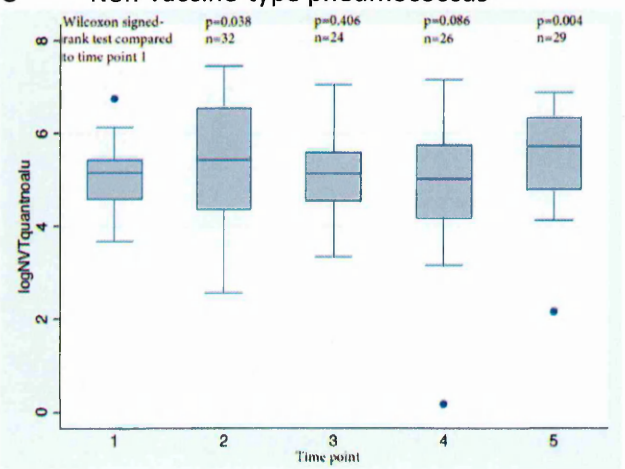


Figure 10.12: Log_{10}-transformed total, vaccine-type and non-vaccine type nasopharyngeal pneumococcal concentration over time. Time point 1 = day 0, time point 2 = day 30, time point 3 = day 60, time point 4 = day 90 and time point 5 = day 180 since vaccination with PCV-10. Data were log_{10}-transformed for display only. P-values were a non-parametric paired analysis of untransformed data. Pneumococcal concentration was in *lytA* copies/mL without adjustment for the concentration of human DNA present in the nasopharyngeal swab transport medium.

10.3.1 Pneumococcal microarray

In three study participants, microarray types generated from the genome backbone were different even when the same serotype apparently persisted in carriage. Specifically, participant 138 had a microarray type 4 / capsular serotype 6A at day 0 and day 30 but this became a microarray type 3 / capsular serotype 6A at day 60 and day 180. Participant 141 likewise had a microarray type 4 / capsular serotype 6A at day 60 and day 90, but a microarray type 9 / capsular serotype 6A at day 180. Finally, participant 107 had a microarray type 1 / capsular serotype 6B at day 30 and a microarray type 2 / capsular serotype 6B at day 60.

10.4 Discussion

Total pneumococcal nasopharyngeal concentration increased in the six months post PCV-10 vaccination. Vaccine-type pneumococcal nasopharyngeal concentration did not change significantly over this time period. Non-vaccine-type pneumococcal

nasopharyngeal concentration increased significantly at the same time. These results held true regardless of whether we looked at only the culture-positive swabs with serotype data by microarray, or whether we looked at all swabs including imputed data for those that were not able to be typed by microarray.

The ratio of median VT concentration to median NVT concentration fell by 69% over the same time period. Taken together, the results of these analyses are compatible with a vaccine effect on nasopharyngeal concentration of vaccine serotypes. It has been clearly demonstrated that pneumococcal conjugate vaccines decrease prevalence of vaccine-type carriage, both in the PRISM study [10] and elsewhere.[156-158] Density of vaccine-type carriage decreased with vaccination in a Navajo population in the USA,[158] and in the Gambia, although in the latter study non-vaccine-type carriage also decreased at the same time.[103]

A major limitation of this study is the lack of microarray data for control children without vaccination, to adequately address confounding by seasonality. Pneumococcal carriage prevalence in Kilifi increases from January to July, with a peak just after the April-July rainy season. [195,204] If microarray data for control children were available, we may have been able to demonstrate a relative decrease in vaccine-type pneumococcal concentration in vaccine recipients compared to controls, and would have been able to quantify the decrease. However, the lack of an increase in vaccine-type pneumococcal concentration when total pneumococcal nasopharyngeal concentration increased with the pneumococcal season is plausibly explained by vaccine effect. We cannot rule out a reason other than seasonality for the increase in total pneumococcal and NVT pneumococcal concentration over the studied time-period, such as a carriage epidemic of NVTs. Studying the total, VT and NVT concentrations over the same time period among

controls would assist in understanding whether seasonality is the explanation and whether VT pneumococcal concentration really does decline in vaccinees relative to controls.

We assessed the differences between paired swabs collected pre-vaccine and swabs collected 6 months later, in VT, NVT and total pneumococcal concentration. It may be more relevant to assess ratios of such paired swabs to detect changes that might affect individual disease risk post-vaccine. We were limited to the assessment of differences rather than ratios by the large number of zero quantities in the data (90 of 169 swabs contained 0 *lytA* copies/μg human DNA of VT pneumococci and 45 of 169 swabs contained 0 *lytA* copies/μg human DNA of NVT pneumococci). Differences may still be relevant; the graphs of untransformed nasopharyngeal pneumococcal concentration by serotype for individual children marked "A" in Appendix C demonstrate that large linear changes in concentration do occur – and these may be important for transmission or disease risk.

We did not have sufficient data to demonstrate changes in individual serotypes over time after PCV-10 vaccination by paired analysis. To investigate whether the vaccine has an effect of reducing the concentration of newly-acquired colonisation with individual vaccine-types, even if it does not entirely prevent acquisition with those types, a larger dataset allowing a paired comparison of serotype-specific concentration would be needed.

Looking at the ratio of VT to NVT pneumococcal concentration gets around the problem of seasonality, as we do not expect season to influence VT or NVT relative concentrations. However, of 65 swabs demonstrating multiple carriage by microarray, only 37 demonstrated multiple carriage with both VT and NVT pneumococci present. These 37

swabs were not significantly different from all other swabs except that they were more likely to come from children with a slightly lower MUAC and from homes with multiple smokers. The ratio of VT:NVT did appear to decrease with time, but numbers were too small to be able to prove statistical significance. In order to increase the number of swabs contributing data for analysis we computed a ratio of median VT to median NVT concentrations using swabs containing VT or NVT. This ratio does not include the swabs with no VT or those with no NVT quantity (for example, a swab with 100% NVT quantity yields a 0% concentration for VT) as seen in tables 10.2 and 10.3 because several median concentrations would have been 0. The log ratio of median concentrations, VT:NVT, fell significantly across time after vaccination with PCV-10. The shape of the curve for VT:NVT ratio across time was similar for the analysis using median concentrations from all carriers as for the restricted analysis using 37 individual swabs with multiple carriage of VT and NVT, seen in figures 10.9 and 10.10.

Another limitation of these data is the 58 swabs that were positive for *lytA* by qPCR but did not yield pneumococcus on culture for microarray serotyping. It is known that the autolysin gene may exist in non-pneumococcal streptococci, such as *Streptococcus pseudopneumoniae*, so we might suspect that these nasopharyngeal swabs were detecting non-pneumococcal species.[108,205] However, as analytical specificity of *lytA* PCR has been demonstrated by several authors, as seen in table 2.1 of the introduction, and as we used the CDC-revised primer and probe sequences [125] and further validated them ourselves on 70 *S. pneumoniae* and 30 non-pneumococcal streptococcal species from Kilifi, we assumed specificity for pneumococcus. The Kilifi data can be seen in section 3.2, chapter three, and further discussion of the specificity of the *lytA* assay for pneumococcus can be found in section 2.2.2 in chapter two. It is also possible that the PCR assay was detecting non-viable pneumococci, since some of the swabs that did not

grow pneumococcus for microarray typing had grown pneumococcus two years previously when the original study was carried out. Imputing data was possible using the ratio of median VT quantity to median NVT quantity among positive swabs from the pneumococcal microarray at each time point; applying this ratio to the concentration of pneumococcus found by *lytA* qPCR for swabs lacking microarray data. Analysing the changes in VT and NVT concentrations across time using the model with imputed data did not change the conclusions drawn from the analysis without imputed data.

Pneumococcal conjugate vaccines are being rolled out to more and more of the world that needs them, and more countries are beginning to monitor impact of the vaccines. Additionally, new-generation pneumococcal vaccines that are not serotype-specific are being developed and tested. Studies using nasopharyngeal carriage as the endpoint instead of IPD are proposed; taking the vaccine effect on carriage to be a marker of direct protection against pneumococcal disease at the individual level.[64] Pneumocarr, a Gates-funded initiative to advise on the use of nasopharyngeal pneumococcal carriage as the endpoint for vaccine studies, have suggested that pneumococcal acquisition, duration of carriage and density of carriage are important parameters to consider.[164]

If the vaccine effect on carriage includes reduction in density of carriage, this will be important to detect. The degree of importance will depend on the relationship between carriage density and transmission and between carriage density and invasion, which we do not yet fully understand. The relevance of duration of carriage for invasive disease depends on whether we believe that invasion occurs only within a month of new acquisition of carriage, but the relevance of duration of carriage for transmission is clear.

Methods that detect multiple serotype carriage are also important to get the full picture.[171] Serotype replacement in carriage can be unmasking or true replacement. Our

data illustrate the use of nasopharyngeal pneumococcal concentration in addition to detection of multiple serotype carriage pre and post vaccination. The graphs by individual study participant in figure 10.4 and Appendix C demonstrate the amazing complexity of pneumococcal dynamics by serotype post vaccine; multiple serotype carriage, dominance of the prevailing type, variation in serotype-specific concentration both within individuals over time and between individuals. These analyses also reveal the inherent difficulties in studying the density of carriage. Paired analysis of serotype-specific nasopharyngeal pneumococcal concentration in swabs before and after vaccine is constrained by the zero concentrations that are present in such data. Restricting the analysis to the small subsets of swabs that are positive for specific serotypes, or to swabs that are positive for either VT or NVT alone, constrains the power of the analyses and ignores the information contained in negative swabs. Negative swabs can contain important information as part of a pair, such as the acquisition of a VT where there was no VT previously, or the loss of a VT where there was previously a VT present.

The pneumococcal microarray, a powerful tool for demonstrating multiple carriage, relative abundance, and genetic relatedness of pneumococci, is currently limited to swabs that are culture-positive with pneumococci. The requirement for a culture step in the standard microarray workflow is a major limitation of the methodology both because it creates an artificial step in the data, where swabs with a lower pneumococcal concentration are less likely to grow and be included in the microarray, and because we cannot be sure that the relative concentrations of bacteria detected in the microarray are the same as they were in the original swab. It is possible to extract DNA directly from the nasopharyngeal swab specimen, and amplify it for use in the microarray, but this method has been shown to be less sensitive for detecting minor serotypes and multiple carriage than use of the culture amplification step (personal communication, Jason Hinds).

Recently, serotype-specific quantitative PCRs have been developed for many pneumococcal serotypes,[142,167] but there will still be swabs that are negative for any given serotype under study. Even when aggregating serotype-specific data into VT and NVT there are still swabs that are negative for all vaccine-types or for all non-vaccine types, so the issue of dealing with quantitative data with medians at or close to zero remains. Serotype-specific PCRs performed directly from nasopharyngeal or oropharyngeal swabs are also troubled by problems with lack of specificity for pneumococcus, as non-pneumococcal streptococci may harbour pneumococcal capsular genes.[167,177]

The nasopharyngeal swabs in this study had a very high prevalence of pneumococcal carriage as determined by *lytA* qPCR. Aliquots of uninoculated STGG were used as negative extraction controls and taken through all the lab procedures, remaining negative. As discussed, we assume the CDC's *lytA* assay that we used is specific for pneumococcus. This leads to the question; what is pneumococcal carriage? In the mid-last century, this was the ability of a nasal aspirate or saliva to grow pneumococcus in the peritoneum of a mouse, more recently it has been growth of pneumococci on blood agar with gentamicin from a nasopharyngeal swab, the WHO standard method.[168]

Austrian recognised that very different surveillance results ensued from different methods of culture for pneumococcus.[23] Austrian also stated that Gundel found in 1933 that "60% of humans could be shown to carry pneumococci if examined appropriately and that, if studied over time, this fraction would approach 100%."[23] Swab positivity in our study of 12-23 month old infants over 6 months was 75% of swabs by culture and microarray, yet 97% of swabs tested positive by *lytA* qPCR. We expect that PCR might be more sensitive than culture, because of the ability to detect non-viable organisms. We may be detecting pneumococcus at levels that do not constitute carriage as we know it.

Most pneumococcal carriage data from around the world is culture prevalence, using methods similar to the WHO standard method.[168] There are some data available using *lytA* qPCR to compare with ours; a study of Peruvian children <35 months of age found carriage among 60.1% by culture compared to 77.4% by PCR, but they used a different cut-off for PCR-positive than we did, setting a Ct value of <35 of 40 cycles as positive whereas we took all amplification curves up to cycle 40 as positive.[176] It can be more difficult to call positives from amplification curves occurring late in the amplification cycles of real-time PCR, because it is more difficult to see whether true exponential amplification is taking place. Late positive curves may also be occurring close to or below the reliable lower limit of detection or linearity, limiting accurate quantification. However, calling all such curves negative may be ignoring true low positives. Various strategies such as performing PCR replicates and calling positives only on samples that are positive from 2/2 or 2/3 replicates may be employed but low-level positives below the limit of detection will stochastically test as negative sometimes – and likewise some negatives may 'become' low positives if repeat testing is employed.

Among 17-month old infants in Fiji vaccinated with PCV-7, carriage prevalence was 57%, but 69% in indigenous Fijians, by *lytA* PCR.[206] This group used a cut-off of 5×10^3 genome-copies/mL for positives, and tested samples in duplicate, requiring both results to be positive for a sample to be called positive. In Dutch primary school aged children, older than our study population, 88% were found to be carrying pneumococcus in saliva, by culture-enriched PCR.[167] Among Nigerian infants <9 months of age assessed prior to introduction of the pneumococcal conjugate vaccine, 90% of infants were carrying pneumococcus by culture.[207] Likewise in The Gambia, by culture, 90% of 11-month old infants carried pneumococcus, and 100% carried pneumococcus at some point during the 12-month study.[208]

Our swab positivity rate was 75% by culture for microarray, varying from 59% to 87% of swabs at any time point, and 97% by *lytA* PCR, varying from 91% to 100% by time point. This may be similar to prevalence among infants in The Gambia or Nigeria, but it would be interesting to know what the PCR-positivity would have been among those populations. In fact, even the culture for microarray gave slightly different results in this study than the original culture done by the WHO standard method two years previously. There were 29 discordant swabs; the 11 that grew originally but failed to do so for this study may have not survived freeze-thaw cycles and 18 swabs yielded pneumococcus by culture for microarray but failed to do so originally by the standard WHO method. These may have been due to the WHO standard method of selecting a single colony for identification and serotyping; it would be possible to miss a pneumococcal colony with that technique compared to the DNA extraction from a sweep suspension of all growth on the plate for microarray. Previous work in our laboratory demonstrated 98% sensitivity of pneumococcal culture from frozen compared to fresh nasopharyngeal swabs in STGG, but swabs were only frozen for 2 months, compared to three years in the current study.[209]

The microarray gives us additional data to consider. The genome backbone of the pneumococci studied can be used to examine the relatedness of the isolates. We can see that some serotypes that appear to persist over time in an individual, are in fact not the same clone, so that replacement has occurred with the same serotype. Such information would not be available from a traditional culture-based study, unless whole genome-sequencing was applied to the cultured isolates, nor from a serotype-specific qPCR-based study.

In summary, vaccine-type pneumococcal nasopharyngeal concentration did not change over six months post PCV-10 among Kenyan children aged 12-23 months. At the same

time, non-vaccine-type and total pneumococcal concentration increased in line with possible seasonal changes in pneumococcal carriage prevalence. This is compatible with a relative effect of vaccine on VT compared to NVT concentration, but controls are needed to confirm and quantify the effect. Such control samples were collected in the parent PRISM study and permission is being sought to extend the microarray analysis to these samples. If there were no effect of vaccine, then we would expect to see similar changes over time in the control arm as in the vaccine arm reported here, ie. no change in VT concentration, and an increase in total pneumococcal and NVT concentration over the six-month period.

The ratio of median VT to median NVT pneumococcal concentration decreased by 69% over the six months post-vaccine. This is likely to have been a vaccine effect on VT relative to NVT concentration.

Colonisation density and the effects of colonisation density on transmission and invasion may vary by serotype, yet large datasets are needed to examine effects by individual serotype. Simplifying serotype-specific data into VT and NVT still leads to complications in analysis, with many zero values and median values at or close to zero; prevalence contributes significantly towards quantitative data. The data are not normal but log-transformation leads to loss of true zero concentrations, so non-parametric analyses are needed.

Carriage defined by pneumococcal real-time PCR cannot be directly compared with carriage defined by standard culture. Measuring serotype-specific density of colonisation for vaccine studies will be challenging and a lot of thought needs to go into planning analyses for such studies. Cross-sectional surveys will be easier to analyse than longitudinal studies using quantitative microarray data but paired analysis of longitudinal

data is epidemiologically stronger. We need to continue to innovate ways to describe pneumococcal dynamics in the nasopharynx.

Chapter Eleven

11 Objective four: Pneumococcal DNA quantification in blood as a diagnostic method for invasive pneumococcal disease

11.1 Introduction

Pneumococcal disease is the leading cause of childhood vaccine-preventable death in the world.[11] Diagnosis of invasive pneumococcal disease (IPD) for either epidemiological studies or clinical care is difficult; a minority of cases are detected by blood culture even under ideal conditions;[12] sputum is not spontaneously produced by young children; it is problematic to differentiate carriage from disease using culture of nasopharyngeal samples; diagnostic serology is insensitive in children and in any case needs paired sampling;[13] and urinary antigen tests are not sufficiently specific for use in young children.[194] It is presumed this lack of specificity of urinary antigen tests among children is due to nasopharyngeal pneumococcal carriage; swallowing of nasopharyngeal secretions, absorption of pneumococcal polysaccharide in the gut and excretion in urine [15] - and perhaps intermittent transient benign bacteraemia.[14]

Pneumococcal nasopharyngeal colonisation is a prerequisite for pneumonia and IPD.[64-67] Recent acquisition of an invasive strain of pneumococcus is thought to be the event that leads to invasive disease if immune mechanisms fail.[66 67] This is based largely on a longitudinal study by Gray *et al*, where in fact only three of the 31 pneumococcal infections that occurred were IPD, the remainder being otitis media. Three-quarters of

the pneumococcal infections in the study by Gray *et al* occurred within a month of acquisition of a new strain of pneumococcus in the nasopharynx. Young children are not only more likely to carry pneumococcus in the nasopharynx, but they may carry a higher density of pneumococcus than older children or adults.[103] Both the higher prevalence of carriage and a higher density of pneumococcal carriage in the nasopharynx could potentially contribute to children from populations with low socio-economic opportunity being at higher risk for pneumococcal pneumonia and IPD; these could also contribute to a lack of specificity in the diagnosis of IPD.

Progress in molecular methods to diagnose IPD has been slow, with several gene targets of NAATs being non-specific to *S. pneumoniae* due to its close homology with related non-invasive species.[106-125] A review of heterogenous studies of molecular methods to diagnose IPD from blood samples carried out over 1993 to 2009 describes sensitivity of 57-66% and specificity of 88-99%. Specificity depended on what sort of control groups were used.[124] Poor sensitivity has been thought to be due to low specimen volume that can be tested by NAATs, PCR-inhibitors and pneumococcal autolysis and DNA degradation from old specimens or specimens stored at -20°C.[116,122,125,127]

Most pneumococcal PCR assays now use the autolysin gene *lytA*,[125] but assays using the pneumolysin gene *ply* were common initially.[106,127] Other pneumococcal gene targets include the surface adhesion protein *psaA*,[125] a gene of unknown function spn9802,[122] the capsular polysaccharide biosynthesis gene cpsA[136] and the bacterial recombinase gene *recA*.[137] Dagan and colleagues suggested that pneumococcal DNA may be detectable in the bloodstream of healthy individuals who do not develop clinical disease, from pneumococci which directly invaded the blood from the nasopharynx or which entered the bloodstream phagocytosed by lymphoid cells.[106] However, they were using the

pneumolysin gene as a PCR target, that we now know detects not only pneumococcus but also other commensal streptococcal flora. This has meant that the concept of pneumococcal DNA being readily detectable in the bloodstream of young children who remain completely well has not received as much consideration in the scientific community as it might have done. Non-pneumococcal commensal streptococcal flora are not uncommon blood culture contaminants in young children,[210] presumably because infants' skin is frequently in contact with their saliva, but they should otherwise not be expected to be detected in blood samples either.

lytA has been shown to be a specific gene target for identifying pneumococcus for PCR assays,[116,125,136,138,139] and although the autolysin gene has been found in some *S. pseudopneumoniae* [128] and *S. mitis* genomes,[108] it can be differentiated from that in *S. pneumoniae*.[128,138] Experience with the *lytA* target for real-time PCR assays in the diagnosis of IPD from blood samples is increasing,[15,74,121,141-145] particularly with the CDC's *lytA* assay published in 2007.[125] Further discussion of the analytical specificity of this assay can be found in section 2.2.2 of chapter two. The sensitivity of *lytA* real-time PCR is now probably better than blood culture, the traditional gold standard for the diagnosis of IPD, with attention to specimen management and use of modern DNA extraction methods.[74,143,144] Clinical specificity has been harder to measure, particularly among children, the population who most need an improved diagnostic assay. In Italy *lytA* real-time PCR from blood was negative among 147 well children with a median age of 4.9 years, half of whom were colonised with pneumococcus in the nasopharynx by PCR of nasopharyngeal swabs. [15] In South Africa real-time *lytA* PCR from serum was negative in 100 well children with a mean age of 5.8 years, half of whom were colonised with pneumococcus in the nasopharynx by culture.[148] In Slovenia, *lytA* real-time PCR was positive from plasma among two (7%) of 29 children with a non-pneumonia acute febrile

illness for which a non-pneumococcal cause was identified, with a mean age of 1.5 years.[146] Both the children who tested positive were colonised with pneumococcus. As blood cultures are themselves not a sensitive gold-standard assay with which to define IPD, it is not surprising that real-time PCR could detect cases that are blood culture negative. This does make it difficult to assess specificity of PCR in the diagnosis of IPD because it is not easy to define a suitable group of true negatives among children who are actually the target population for the diagnostic assay – ie. sick children who might have IPD. The use of healthy children or adults as control populations is likely to produce a biased estimate of clinical specificity.

Quantitative real-time pneumococcal PCR has been used from respiratory specimens to separate colonisation from disease [126] and from blood as a marker of disease severity.[74,145,147,149] Quantitative PCR (qPCR) using the *lytA* target on blood samples might help us separate children with IPD from children with PCR-positivity in blood due to nasopharyngeal pneumococcus advancing into the bloodstream without resulting in sepsis, if such a phenomenon really occurs.

11.1.1 Objective

To investigate the test parameters of *lytA* qPCR on paediatric blood specimens in the diagnosis of invasive pneumococcal disease in diverse populations in the developing world.

11.2 Methods

There were two sets of samples available to assess the test parameters of *lytA* qPCR on blood as a diagnostic test for IPD. The first was the Kilifi set of samples, all paediatric medical admissions to Kilifi County Hospital from September 2010 until November 2013;

and two control groups consisting of healthy children and children with upper respiratory tract infection symptoms in the community.[5]

The second was the Pneumonia Etiology Research for Child Health (PERCH) case-control sample set, consisting of children 1 month to 5 years of age enrolled in an international multi-site case-control study of causes of pneumonia. The PERCH sites included Kilifi in Kenya as well as sites in The Gambia, Mali, Zambia, South Africa, Bangladesh and Thailand. Children were enrolled from August 2011 until late 2013 or early 2014, depending on the site.

11.2.1 Kilifi sample set

The Kilifi set of samples came from 8,158 children <13 years of age; 536 community controls from the Kilifi arm of the PERCH study (both with and without symptoms of upper respiratory tract infection), 557 cases of WHO-defined severe or very severe pneumonia admitted to Kilifi County Hospital (KCH) and included in the PERCH study, 6,996 general medical admissions to the paediatric service at KCH, and 69 convenience controls without pneumonia enrolled under non-PERCH studies of pneumonia. The total numbers of paediatric admissions to KCH, the paediatric medical admissions (excluding accidents, surgical cases, a history of poisoning, and fresh burns) and those providing consent for collection of blood for research, can be seen in figure 11.1. Samples were collected from November 2010 until November 2013, but were limited to June 2011 until November 2013 for controls. The 69 convenience controls were collected during the piloting of the PERCH study control selection system in June-August 2011 or were collected during the PERCH study but outside the PERCH age limits so were not included as PERCH data. All controls lived within the Kilifi Health and Demographic Surveillance System (KHDSS) site [196] and were sampled in the community, matched by age and month

of enrolment to cases of WHO-defined severe or very severe pneumonia admitted to KCH. As the pneumococcal conjugate vaccine was introduced to Kenya in January 2011, along with a catch-up campaign for children <5 years of age in Kilifi County, most samples were collected after the introduction of vaccine.

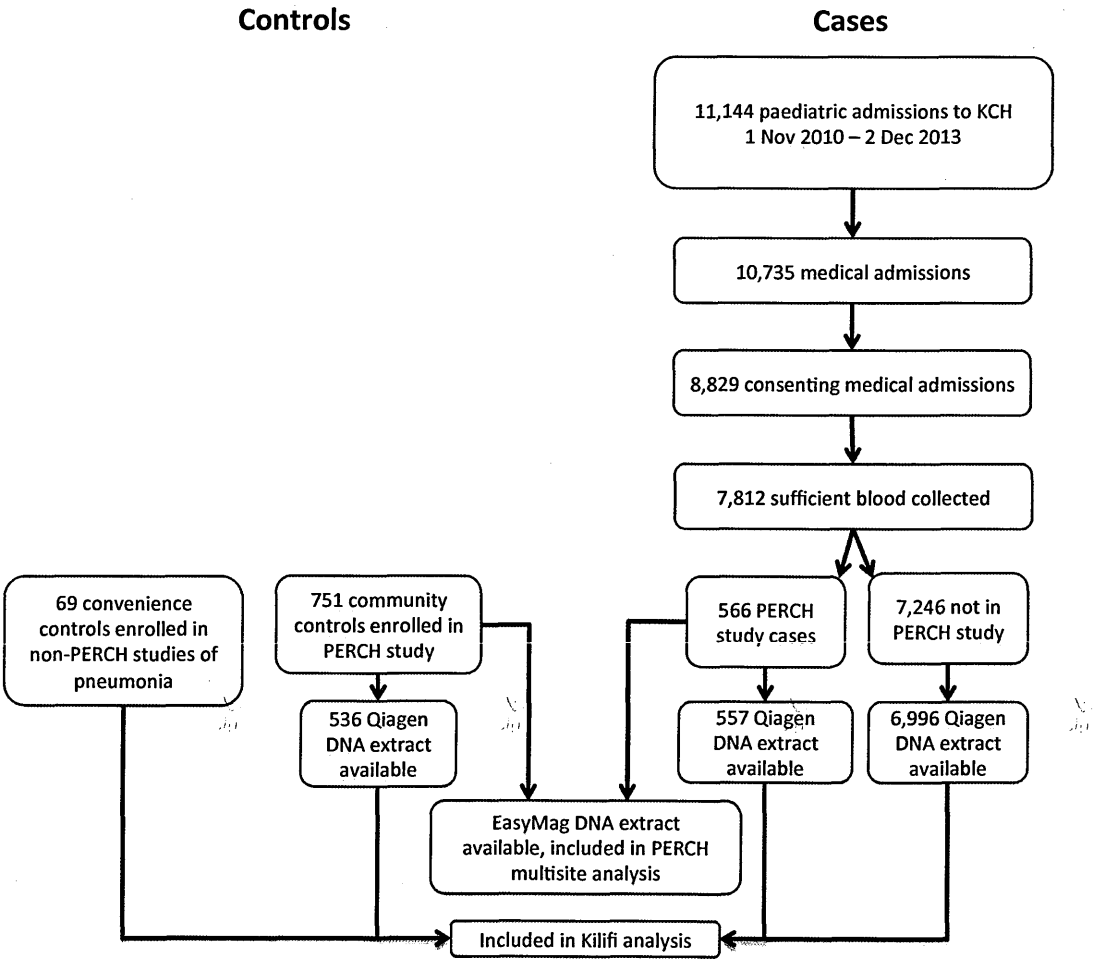


Figure 11.1: Overlap of cases and controls in the two sample sets: paediatric admissions to Kilifi County Hospital (KCH) and Pneumonia Etiology Research for Child Health (PERCH) multi-site study samples, including Kilifi. Samples in the Kilifi analysis underwent DNA extraction by the Qiagen manual method; samples in the PERCH analysis, including Kilifi samples, underwent DNA extraction by the EasyMag automated instrument; samples in both sample sets were extracted by both methods.

Administration of antibiotics prior to sample collection (n=1,260), presence or absence of pneumococcus in the nasopharynx by culture or PCR (n=2,401) and HIV status (n=5,855) were available for some participants. Numbers with available data varied because clinical and laboratory parameters were introduced in a stepwise fashion as the PERCH study was developed and incorporated into existing Kilifi pneumonia surveillance.[5] Additionally, children otherwise meeting the PERCH inclusion criteria of WHO-defined severe or very severe pneumonia, but who were outside the PERCH study age limits (1-59 months) or catchment area (the KHDSS) were included in Kilifi pneumonia surveillance. Administration of antibiotics in hospital was recorded and a serum bioassay was performed during pneumonia aetiology studies. Details of the serum antibiotic bioassay may be found in section 7.3 of chapter seven. Antibiotics were considered as having been administered prior to sample collection if the clinical records indicated this or if the serum antibiotic bioassay was positive.

A whole blood specimen was collected into EDTA from consenting medical paediatric admissions alongside a blood culture and standard admission blood tests. Controls had an EDTA blood sample collected but not a blood culture. The EDTA blood sample was collected into a dedicated tube for DNA extraction, with the aim of extracting DNA within 72 hours of sample collection, and samples were kept refrigerated at 2-8°C until extraction. All samples were extracted using the QiaAmp DNA blood mini kit from Qiagen (section 7.6.1 in chapter seven). Extracted DNA was frozen at -80°C until undergoing quantitative *lytA* PCR as described in section 7.9 of chapter seven.

11.2.2 PERCH sample set

The PERCH sample set consisted of 8,981 children 1-59 months old, cases being those admitted to hospital with symptoms meeting the WHO syndromic definition of severe or

very severe pneumonia, and age and time-matched community controls without pneumonia at each of the seven sites.[4] The WHO definition for severe pneumonia is cough or difficulty breathing, with lower chest wall indrawing. Very severe pneumonia is defined as severe pneumonia plus any of; central cyanosis, difficulty in breastfeeding/drinking or vomiting everything, convulsions/lethargy/unconsciousness or head nodding. The cases and controls from Kilifi that came from within the KHDSS, were 1-59 months of age and fell between 15th August 2011 and 15th November 2013, were also included in the PERCH all-site dataset. Pneumococcal conjugate vaccine (PCV) had been introduced as PCV-10 in Kenya in January 2011, PCV-7 in the Gambia in August 2009, PCV-13 in the Gambia in April 2011, PCV-13 in Mali in March 2011, PCV-7 in April 2009 in South Africa, PCV-13 in June 2011 in South Africa, and PCV-10 in Zambia in July 2013. Thailand and Bangladesh had not introduced pneumococcal conjugate vaccines. All PERCH samples were extracted using the EasyMag automated extraction instrument (section 7.7, chapter seven). PERCH samples extracted by EasyMag were done in batches for logistical reasons and the time from sample collection to DNA extraction varied considerably.

The subset of PERCH samples that were from Kilifi had therefore been extracted by two methods, the QIAamp DNA blood mini kit manual spin-column method and the EasyMag automated extraction instrument. Extracted DNA from whole blood samples underwent *lytA* qPCR (section 7.9, chapter seven). All Ct values <40 were considered positive, as long as an exponential amplification curve was present. Results of *lytA* qPCR from the Qiagen extraction are included in the Kilifi dataset and results of *lytA* qPCR from the EasyMag extraction are included in the PERCH all-site dataset.

The samples from all other PERCH sites followed the same protocols for EasyMag extraction and *lytA* qPCR as those in Kilifi, but were run in the individual site laboratories, with standardised quality control procedures and external quality assurance.

11.2.2.1 Nasopharyngeal culture and PCR in PERCH

Nasopharyngeal flocked swabs in STGG (NP-STGG) transport medium underwent broth enrichment and culture on blood agar plates containing gentamicin to select for pneumococcus. Further details of broth enrichment culture can be found in section 7.1 of chapter seven.

Nasopharyngeal flocked swabs and oropharyngeal swabs were collected and combined in Copan universal transport medium (NP/OP-UTM), underwent extraction using the EasyMag automated instrument, and multiplex PCR using the Fast-Track Diagnostics respiratory 33-plex quantitative real-time PCR, which includes a *lytA* target (Fast-Track Diagnostics, Luxembourg).

11.2.2.2 Blood cultures

Blood cultures were performed using the BACTEC 9050 automated blood culture system in Peds Plus bottles from BACTEC (Becton Dickinson, NJ, USA). Positive blood cultures were gram-stained, sub-cultured and organisms were identified using standard microbiological procedures.

11.2.3 Analysis

The analysis of the Kilifi dataset informed the analysis of the PERCH all-site data.

11.2.3.1 Kilifi data analysis

Characteristics of cases/general admissions and controls in the Kilifi sample set were compared using a Wilcoxon rank-sum test or chi-squared test as appropriate.

lytA qPCR data was categorised as a binary variable, either positive or negative, and compared to blood cultures as the gold standard, to assess the sensitivity of *lytA* qPCR in the diagnosis of IPD. Specificity was assessed among community controls, who were well or had symptoms of upper respiratory tract infection, so were very unlikely to have IPD.

A quantitative comparison of pneumococcal concentration in whole blood between participants admitted to hospital and community controls was carried out. *lytA* qPCR results on whole blood were examined by presence or absence of pneumococcus in the nasopharynx as detected by culture or multiplex PCR.

As the *lytA* qPCR assay was found to have poor specificity in the diagnosis of IPD, further investigation of assay performance was carried out. A receiver operating characteristic (ROC) curve was created for Ct values from the *lytA* qPCR, using IPD cases according to blood culture to set the true positives and community controls to set the true negatives. The ROC curve was used to set a cut-off to optimise sensitivity and specificity of *lytA* qPCR for the diagnosis of IPD. This cut-off was used to create categories of *lytA* qPCR results; negative, weak-positive and strong-positive. Weak-positive results were positive results with a Ct value above the cut-off, strong-positive results were positive results with a Ct value below the cut-off. Weak and strong-positive *lytA* qPCR results were then analysed by sample collection date, extraction run, PCR run and laboratory technician.

Univariable and multivariable logistic regression were used to assess variables associated with weak positives or strong positives by *lytA* qPCR. Variables associated with outcome

at a p-value of <0.1 in univariable analysis, or that were considered important for the model, were included in multivariable analysis models by backwards stepwise selection. Biologically plausible interactions were assessed using the likelihood ratio test. Weak-positive samples were analysed relative to negative samples and strong positive samples were analysed relative to weak-positives and negatives combined, assuming that weak positives may have been false positive samples.

The ROC curve was then re-created in the subset of Kilifi participants less than five years of age, using the raw *lytA* quantity from the qPCR, unadjusted for extraction:elution ratio, as this is standardised across PCR runs, unlike the Ct value which may vary run to run. Potential cut-offs were assessed from this Kilifi ROC curve to apply to the PERCH all-site dataset.

The *lytA* concentration values from whole blood samples that had been extracted by both the QIAamp DNA blood mini kit and the EasyMag instrument, were $\log_{10}$-transformed and plotted in a scatterplot and a z-test for binary data was performed. For illustrative purposes only, negative values were assigned a value of 1 prior to log transformation for the scatterplot.

11.2.3.2 PERCH data analysis

Analysis of the PERCH dataset began by assessing between-site differences in positivity and concentration of *lytA* from whole blood. Descriptive analyses were performed: the whole blood *lytA* concentration by site and case-control status; participant characteristics by case-control status with chi-squared test for differences between cases and controls; and the distribution of *lytA* concentration by participant characteristics with the Wilcoxon rank-sum test for binary data and the Kruskal-Wallis test for categorical data of more than

two categories. The correlation between nasopharyngeal pneumococcal concentration and whole blood pneumococcal concentration for positive samples was assessed using Pearson's pairwise correlation co-efficient and displayed as a scatterplot with a line of best fit by linear regression.

Two potential cut-offs from the Kilifi under-five years of age dataset were selected from the ROC plot based on maximising sensitivity and specificity. One of the cut-offs was also noted as being close to the lower limit of detection and linearity of the assay. This cut-off was used to define weak-positive and strong-positive samples for *lytA* qPCR in the PERCH all-site dataset.

To explore associations with strong positivity or weak positivity univariable logistic regression was performed for weak-positive outcome and strong-positive outcome and variables selected based on a p-value <0.1, biological importance for the model, and avoidance of collinearity. Multivariable logistic regression was performed by backwards stepwise selection, for weak-positive and strong-positive outcome. Strong-positive samples were compared with weak-positive samples and negative samples combined; weak-positive samples were compared with negative samples only. For each model, biologically plausible interaction terms with a significance level of <0.05 by the likelihood ratio test were included in the model. Whenever observations were dropped from the model due to perfect prediction, Firth logistic regression was employed.

The sensitivity and specificity of a strong positive outcome were calculated for each site in the PERCH dataset and these were plotted in a ROC plot with 95% confidence intervals. A chi-squared or Fisher's exact test, as appropriate, was used to assess heterogeneity in sensitivity and specificity between sites.

The number of extra diagnoses of IPD that could be made in the PERCH study by employing the cut-off for *lytA* qPCR, in addition to those diagnoses made by blood culture, was calculated.

11.2.4 Ethical Clearance

The study was approved by the Kenyan National Ethical Review Committee (SSC 1932). In the original studies consent was obtained from guardians of all children studied.

11.3 Results

11.3.1 Kilifi results

11.3.1.1 Characteristics of cases and controls in Kilifi

There were 8,158 samples with *lytA* qPCR results in the Kilifi dataset, among children from the first day of life to 13 years of age. Age, HIV positivity, days from specimen collection until DNA extraction, antibiotic use prior to sample collection, nasopharyngeal carriage of pneumococcus as detected by culture of NP-STGG, and carriage as detected by culture of NP-STGG or multiplex PCR of NP/OP-UTM, and blood culture positivity for *S. pneumoniae* can be seen in table 11.1. The 63 children that had IPD as defined by a blood culture that grew *S. pneumoniae*, had a median age of 2.8 years (IQR 1.1 – 4.4 years, range 3 days – 13.5 years).

Table 11.1: Characteristics of general admissions and controls in the Kilifi sample set

	Controls			Cases/General admissions			
	Median	IQR		Median	IQR		p-value*
Age (months)	13.2	6	25.2	15.6	3.4	46.8	0.355
Days from specimen collection to DNA extraction§	2	1	3	2	1	3	<0.001
	n	N	(%)	n	N	(%)	
Female	284	605	(47)	3,182	7,553	(42)	0.021
HIV positivity	2	478	(0.4)	407	6,786	(6.0)	<0.001
Antibiotics prior to sample collection	13	551	(2.4)	270	1,068	(25.3)	<0.001
Carriage of pneumococcus by culture	470	600	(78)	953	1,879	(51)	<0.001
Carriage of pneumococcus by culture or PCR	528	595	(89)	1,324	1,806	(73)	<0.001
Blood culture positive for *S. pneumoniae*	-	-	-	63	7,396	(0.9)	-

*Wilcoxon rank sum for comparison of medians, chi-squared test for comparison of proportions

§the distribution of days from specimen collection to DNA extraction was slightly longer for cases than for controls

11.3.1.2 *lytA* qPCR positivity among cases and controls in Kilifi

Of 605 controls, 31 (5.1%) were *lytA* qPCR positive from whole blood and of 7,553 PERCH cases and general admissions, 305 (4.0%) were *lytA* qPCR positive. The breakdown by participant group can be seen in table 11.2. The differences between groups were not significant, with a chi-squared p-value of 0.398. Specificity of *lytA* qPCR was therefore estimated at 94.9% among the 605 controls. Sensitivity of *lytA* qPCR estimated against IPD as defined by blood culture positivity among cases and general paediatric admissions was 39/63 (62%).

Table 11.2: *lytA* PCR positivity from blood samples by participant group in the Kilifi sample set. PERCH = Pneumonia Etiology Research for Child Health.

	lytA positive	(%)	Total
PERCH controls	29	5.4	536
PERCH cases	25	4.5	557
General admissions	280	4.0	6,996
non-PERCH controls	2	2.9	69
Total	336	4.1	8,158

11.3.1.3 *lytA* qPCR quantitative data in Kilifi

The distribution of log_{10}-transformed *lytA* concentration for all positive samples by participant group, including cases of IPD as defined by blood culture, can be seen in figure 11.2. The median *lytA* concentration among controls with a positive *lytA* qPCR was 2.51 (IQR 2.07-3.21) log_{10} copies/mL and among pneumonia cases and general admissions was 2.33 (IQR 2.02-3.10) log_{10} copies/mL, rank-sum p-value 0.656. The distribution of log_{10} *lytA* concentrations in blood by participant characteristics can be seen in table 11.3. Concentration varied by blood culture positivity as expected, and by DNA extraction period, but without a discernable trend. The whole blood log_{10} *lytA* concentration by nasopharyngeal carriage status, for general admissions and pneumonia cases and for controls, can be seen in figure 11.3.

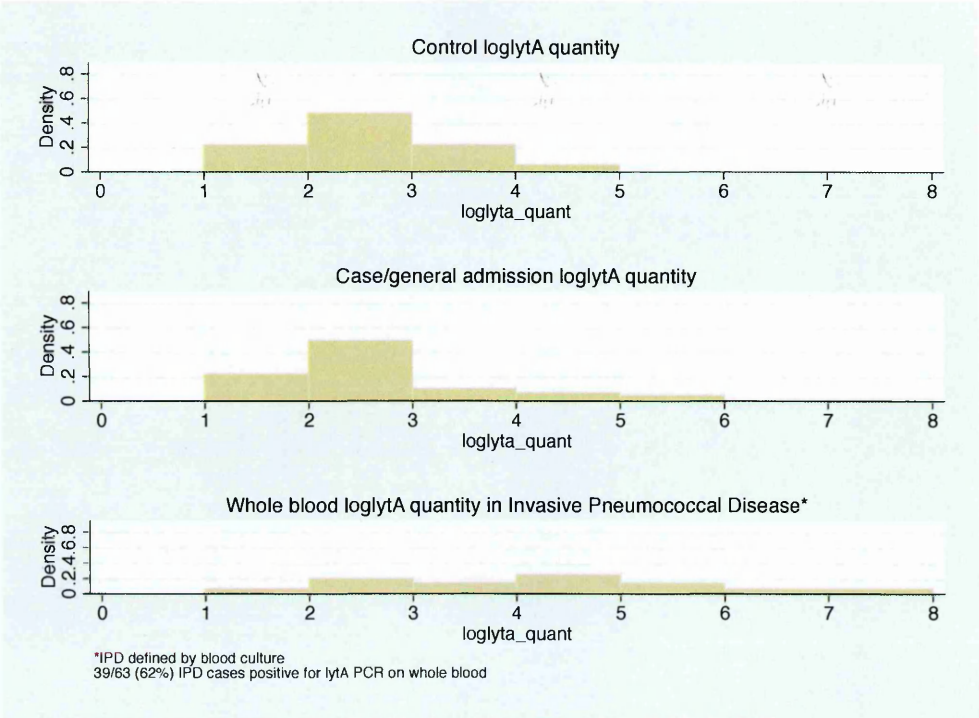


Figure 11.2: Distribution of log_{10}-transformed *lytA* concentration in log_{10} copies/mL, in blood samples in controls, in cases/general admissions and in invasive pneumococcal disease (IPD) in the Kilifi sample set. IPD was defined by growth of *S. pneumoniae* from a blood culture. IPD cases are a subset of cases/general admissions.

Table 11.3: Log_{10}-transformed *lytA* concentration in log_{10} copies/mL, in blood samples, by participant characteristic in the Kilifi sample set.

Characteristic	Number *lytA* positive	Median log(10) *lytA* quantity copies/mL	IQR		p-value*
Gender					0.419
Male	181	2.31	2.00	3.03	
Female	155	2.39	2.07	3.20	
Age in tertiles					0.588
0-7 months	84	2.29	2.07	2.90	
7-31 months	121	2.32	1.99	3.02	
31 months - 13.9 years	131	2.46	2.03	3.22	
Invasive pneumococcal disease					<0.001
Negative	262	2.25	2.00	2.80	
Positive	39	4.23	2.73	5.48	
Nasopharyngeal carriage by culture or PCR					0.645
Negative	7	2.52	2.09	3.10	
Positive	104	2.51	2.12	3.29	
Antibiotics prior to collection of blood sample					0.448
Absent	58	2.52	2.12	3.30	
Present	13	2.74	2.37	3.10	
HIV status					0.319
Negative	269	2.36	2.01	3.09	
Positive	35	2.68	2.04	3.47	
Case-control status					0.194
PERCH community control	29	2.46	2.07	2.86	
non-PERCH control	2	3.71	3.35	4.08	
PERCH case	25	2.52	2.18	3.10	
General admission	280	2.31	2.01	3.11	
Extraction period					0.011
Jul-Dec 2010	32	2.38	1.92	2.80	
Jan-Jun 2011	87	2.38	2.04	3.92	
Jul-Dec 2011	61	2.19	1.96	2.56	
Jan-Jun 2012	53	2.43	2.12	3.70	
Jul-Dec 2012	36	2.21	1.89	2.78	
Jan-Jun 2013	34	2.20	1.96	2.89	
Jul-Dec 2013	33	2.52	2.18	4.29	
Days from specimen collection to DNA extraction					0.202
Within 3 days	278	2.32	2.01	3.09	
More than 3 days	58	2.55	2.04	3.13	

*Wilcoxon rank-sum for binary categorical data, Kruskal-Wallis for >2 categories

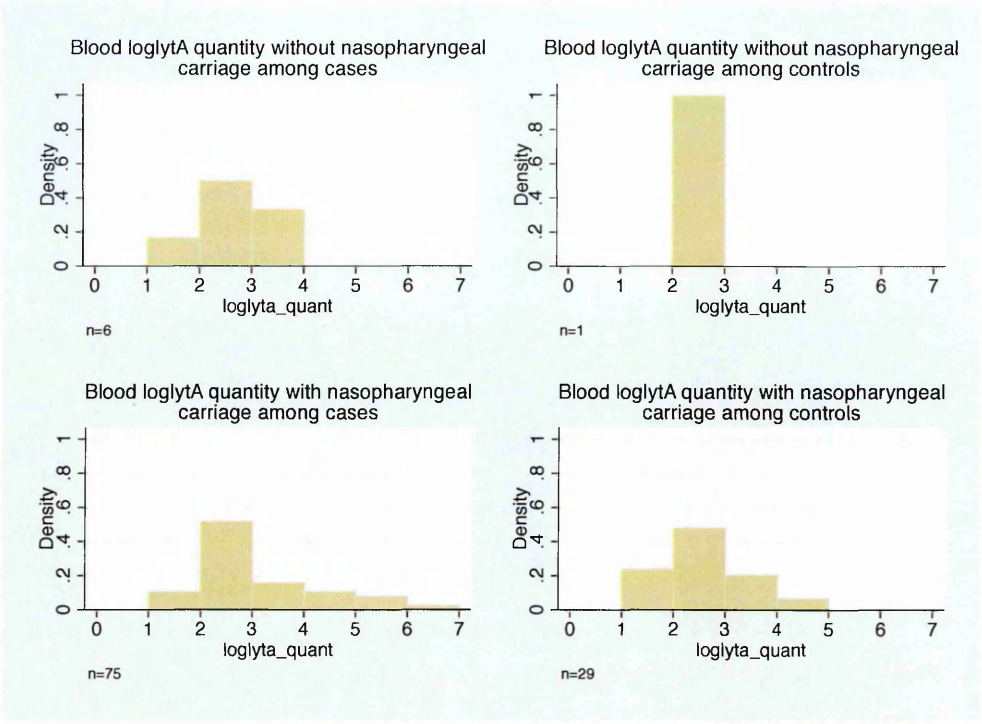


Figure 11.3: Whole blood *lytA* concentration in log_{10} copies/mL by participant group and nasopharyngeal carriage status in the Kilifi sample set. Cases include general paediatric admissions to Kilifi County Hospital. Carriage was defined as culture of *S. pneumoniae* from a nasopharyngeal swab or *lytA* PCR positivity from a nasopharyngeal/oropharyngeal swab.

11.3.1.4 Exploring lytA qPCR assay performance in Kilifi ; categorising positive results

Having seen that specificity of the *lytA* assay was poor, and that there were many low positive results, we set out to investigate the assay performance to try and understand the problem. A receiver operating characteristic (ROC) curve was created using controls to set specificity and cases of IPD to set sensitivity, in order to explore a cut-off to categorise positives into strong and weak positive results. The initial ROC curve can be seen in figure 11.4. From this curve, a Ct value of 38 was selected as a cut-off, with a

sensitivity of 56% and a specificity of 96%. Samples with a Ct value of <38 were categorised as strong positives and samples with a Ct value of ≥38 were categorised as weak positives. Categories of positivity were then explored by sample collection date, DNA extraction date, PCR run number, laboratory operator for DNA extraction, laboratory operator for PCR, to look for laboratory effects. Weak positive *lytA* qPCR results declined over time by DNA extraction period and by PCR period, strong positive *lytA* qPCR results did not decline as much over time by DNA extraction period and did not decline at all by PCR period, as can be seen in table 11.4. Positivity by laboratory operator can be seen in table 11.5 and table 11.6; differences in positivity by operator are confounded by differences over time. There was no evidence of an effect of seasonality when categories of positivity were examined by month of sample collection, as seen in figure 11.5.

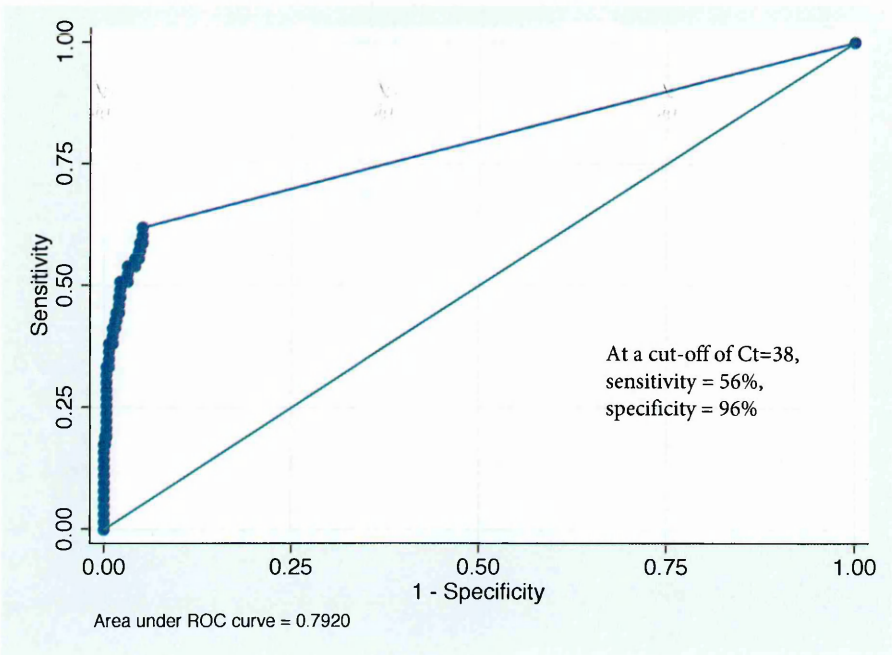


Figure 11.4: Receiver operating characteristic (ROC) curve using controls to set specificity and cases of invasive pneumococcal disease (IPD) to set sensitivity, in the

Kilifi sample set. The ROC curve was used to select a Ct cut-off to categorise positives into strong and weak positive results.

Table 11.4: Whole blood *lytA* qPCR weak and strong positive samples by DNA extraction period and by PCR period during the study in the Kilifi sample set.

Extraction period	*lytA* weak positive	(%)	*lytA* strong positive	(%)	Total
Jul-Dec 2010	20	10.0	12	6.0	200
Jan-Jun 2011	30	1.9	57	3.6	1,574
Jul-Dec 2011	16	1.1	45	3.0	1,479
Jan-Jun 2012	6	0.4	47	3.3	1,432
Jul-Dec 2012	6	0.4	30	2.2	1,363
Jan-Jun 2013	1	0.1	33	3.1	1,065
Jul-Dec 2013	1	0.1	32	3.1	1,045
Total	80	1.0	256	3.1	8,158
Chi-squared p-value compared to negatives		p<0.001		p=0.047	

PCR period	*lytA* weak positive	(%)	*lytA* strong positive	(%)	Total
Jan-Jun 2012	52	3.2	48	2.9	1,650
Jul-Dec 2012	3	1.8	8	4.9	164
Jan-Jun 2013	3	0.2	50	2.9	1,746
Jul-Dec 2013	1	0.1	31	3.0	1,047
Jan-Jun 2014	21	0.6	119	3.4	3,551
Total	80	1.0	256	3.1	8,158
Chi-squared p-value compared to negatives		p<0.001		p=0.566	

Table 11.5: Whole blood *lytA* PCR weak and strong positive samples by laboratory operator and DNA extraction period during the study in the Kilifi sample set.

	Laboratory operator 1					Laboratory operator 2				
	lytA category					*lytA* category				
Extraction	weak		strong		Sub-	weak		strong		Sub-
period	positive	(%)	positive	(%)	total	positive	(%)	positive	(%)	total
Jul-Dec 2010	19	10.8	11	6.3	176	0	0.0	0	0.0	2
Jan-Jun 2011	6	4.0	13	8.6	152					
Jul-Dec 2011	5	3.3	3	2.0	154					
Jan-Jun 2012	0	0.0	0	0.0	26					
Jul-Dec 2012	0	0.0	3	3.6	83	2	0.5	10	2.3	432
Jan-Jun 2013						0	0.0	16	5.1	314
Jul-Dec 2013						1	0.2	8	1.8	438
Sub-totals	30	5.1	30	5.1	591	3	0.3	34	2.9	1,186
Chi-squared					p<0.001					p=0.154

	Laboratory operator 3					Other laboratory operators				
	lytA category					*lytA* category				
Extraction	weak		strong		Sub-	weak		strong		Sub-
period	positive	(%)	positive	(%)	total	positive	(%)	positive	(%)	total
Jul-Dec 2010										
Jan-Jun 2011						0	0.0	1	7.1	14
Jul-Dec 2011										
Jan-Jun 2012										
Jul-Dec 2012	0	0.0	5	2.0	249					
Jan-Jun 2013	0	0.0	8	2.1	379	0	0.0	2	7.7	26
Jul-Dec 2013	0	0.0	8	2.7	294					
Sub-totals	0	0.0	21	2.3	922	0	0.0	3	7.4	40
p-value					p=0.823					NA

Table 11.6: Whole blood *lytA* PCR weak and strong positive samples by laboratory operator and PCR period during the study in the Kilifi sample set

	Laboratory operator 1					Laboratory operator 3				
	lytA category					*lytA* category				
	weak		strong		Sub-	weak		strong		Sub-
PCR period	positive	(%)	positive	(%)	total	positive	(%)	positive	(%)	total
Jan-Jun 2012	52	3.5	48	3.2	1,509					
Jul-Dec 2012	3	1.8	8	4.9	164					
Jan-Jun 2013	2	0.2	25	2.8	892	0	0.0	1	1.0	98
Jul-Dec 2013	0	0.0	9	2.7	334					
Jan-Jun 2014	14	0.6	82	3.8	2,178	5	0.6	30	3.6	843
Sub-totals	71	1.4	172	3.4	5,077	5	0.5	31	3.3	941
Chi-squared p-value					p<0.001					p=0.303

	Laboratory operator 4				
	lytA category				
	weak		strong		Sub-
PCR period	positive	(%)	positive	(%)	total
Jan-Jun 2012	0	0.0	0	0.0	140
Jul-Dec 2012					
Jan-Jun 2013	1	0.1	24	3.2	755
Jul-Dec 2013	1	0.1	22	3.1	713
Jan-Jun 2014	2	0.4	7	1.3	532
Sub-totals	4	0.2	53	2.5	2,140
Chi-squared p-value					p=0.101

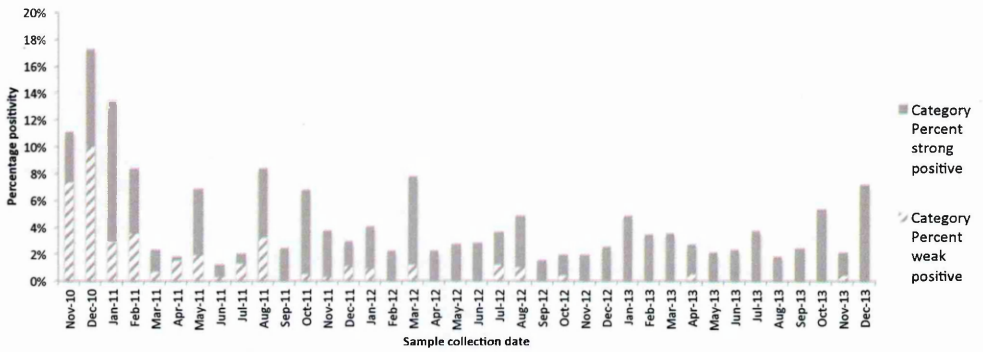


Figure 11.5: Percent positivity of *lytA* PCR on whole blood specimens in the Kilifi sample set by month of sample collection. Categories of positivity are defined as weak positives with a Ct ≥38 and strong positives with a Ct<38.

Univariable and multivariable logistic regression analyses can be seen in table 11.7 for strong positive *lytA* qPCR and table 11.8 for weak positive *lytA* qPCR as outcomes. Variables significant at a level of $p<0.1$ in either of the models, avoiding collinearity, and participant status (whether a control or a case/general admission), were carried through into the multivariable models.

Table 11.7: Univariable and multivariable logistic regression for associations with a strong positive *lytA* PCR on blood, in the Kilifi sample set. A Ct value of <38 was used as a cut-off to define strong positivity. Nasopharyngeal carriage was defined as culture or PCR positivity.

Strong-positive outcome	Univariable analysis				Multivariable analysis				
	n	OR	95% CI		p-value	OR	95% CI		p-value
Nasopharyngeal carriage	2401	4.00	1.84	8.68	<0.001	4.14	1.79	9.57	0.001
HIV	7264	2.26	1.49	3.42	<0.001	2.52	1.33	4.78	0.005
Antibiotic prior to sample collection	1619	1.26	0.68	2.36	0.462				
Age relative to 0-7 months:									
7-31 months	8158	1.47	1.07	2.03	0.019	1.06	0.63	1.79	0.818
31 months - 13.9 years	8158	1.57	1.14	2.16	0.006	1.53	0.87	2.70	0.143
Days from sample collection to DNA extraction	8158	1.00	1.00	1.01	0.137				
Gender	8158	0.92	0.71	1.18	0.501				
Case-control status	8158	0.73	0.48	1.12	0.146				
DNA extraction period	8158	0.94	0.89	1.00	0.060				
DNA extraction period relative to Jul-Dec 2010:									
Jan-Jun 2011		0.59	0.31	1.12	0.105				
Jul-Dec 2011		0.49	0.26	0.95	0.034				
Jan-Jun 2012		0.53	0.28	1.02	0.058				
Jul-Dec 2012		0.35	0.18	0.70	0.003				
Jan-Jun 2013		0.50	0.25	0.99	0.046				
Jul-Dec 2013		0.49	0.25	0.98	0.043				

Table 11.8: Univariable and multivariable logistic regression for associations with a weak positive *lytA* PCR on blood, in the Kilifi sample set. A Ct value of ≤38 was used as a cut-off to define weak positivity. Nasopharyngeal carriage was defined as culture or PCR positivity.

Weak-positive outcome	Univariable analysis					Multivariable analysis			
	n	OR	95% CI		p-value	OR	95% CI		p-value
Nasopharyngeal carriage	2303	8.38	0.50	141.15	0.140				
HIV	7029	2.31	1.10	4.86	0.028				
Age relative to 0-7 months:									
7-31 months	7902	1.42	0.80	2.53	0.231	1.35	0.75	2.41	0.315
31 months - 13.9 years	7902	1.63	0.93	2.86	0.087	1.71	0.97	3.01	0.063
Days from sample collection to DNA extraction	7902	1.00	1.00	1.01	0.090				
Case-control status	7902	0.98	0.42	2.25	0.956	0.36	0.15	0.87	0.024
Gender	7902	0.70	0.45	1.08	0.110				
Antibiotic prior to sample collection	1557	0.25	0.01	4.29	0.338				
DNA extraction period	7902	0.55	0.48	0.63	<0.001	0.53	0.46	0.61	<0.001
DNA extraction period relative to Jul-Dec 2010:									
Jan-Jun 2011		0.17	0.09	0.31	<0.001				
Jul-Dec 2011		0.09	0.05	0.19	<0.001				
Jan-Jun 2012		0.04	0.01	0.09	<0.001				
Jul-Dec 2012		0.04	0.02	0.10	<0.001				
Jan-Jun 2013		0.01	0.00	0.06	<0.001				
Jul-Dec 2013		0.01	0.00	0.06	<0.001				

Nasopharyngeal carriage and HIV positivity were associated with strong positive *lytA* qPCR. Control status and extraction period were associated with a weak positive *lytA* qPCR and it was considered that weak positive results could be false positives. The analysis of the Kilifi data informed the analysis of the PERCH all-site data, seen next. The ROC plot for the Kilifi dataset was re-drawn, using raw quantity from the *lytA* qPCR (unadjusted for extraction:elution ratio), after limiting the dataset to children < 5 years of age (n=6,691) so that it would more closely approximate the PERCH age distribution. Two potential cut-offs were selected based on optimising sensitivity and specificity and were applied to the PERCH all-site data. These can be seen in figure 11.6.

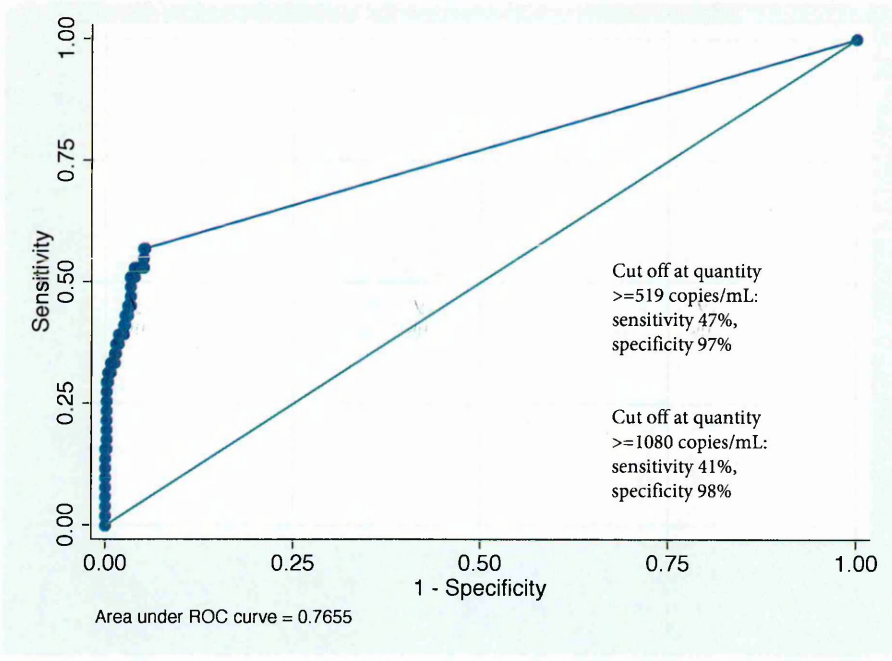


Figure 11.6: Receiver operating characteristic (ROC) curve using controls to set specificity and cases of invasive pneumococcal disease (IPD) to set sensitivity, in the Kilifi sample set limited to children <5 years of age. The ROC curve was used to select a quantitative cut-off to categorise positives into strong and weak positive results to apply to the PERCH multi-site sample set.

11.3.1.5 Comparison of Qiagen and EasyMag DNA extraction methods for *lytA* qPCR on blood, in Kilifi

There were 1,194 samples that underwent both Qiagen extraction and *lytA* qPCR for inclusion in the larger Kilifi dataset and EasyMag extraction and *lytA* qPCR for inclusion in the PERCH dataset. Of these samples, 1,101 were negative by both methods, 29 were positive by both methods, 39 were positive from the EasyMag extract but negative from the Qiagen extract and 25 were positive from the Qiagen extract but negative from the EasyMag extract. A z-test yielded a p-value of 0.271, indicating that there was not a significant bias towards one method or the other. Replacing negative results as a quantity of 1 before $\log_{10}$-transformation revealed the scatterplot found in figure 11.7.

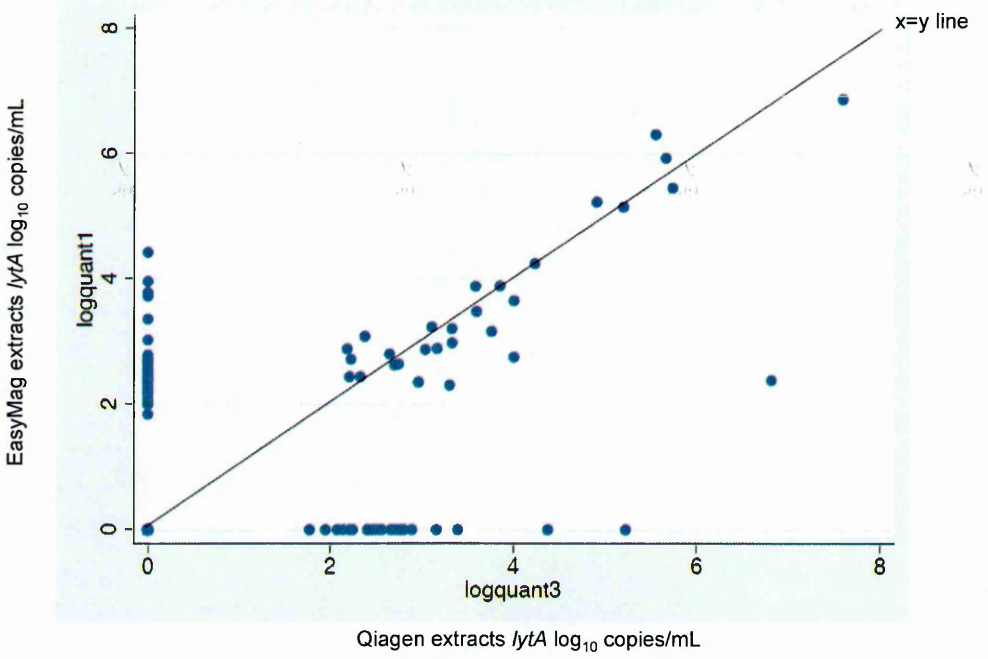


Figure 11.7: Pneumococcal concentration by *lytA* qPCR in $\log_{10}$ copies/mL from paired whole blood samples that underwent DNA extraction by the Qiagen manual method and by the EasyMag automated method, from children in the Kilifi sample set.

11.3.2 PERCH all-site results

11.3.2.1 *lytA* qPCR positivity among cases and controls in PERCH

The PERCH multi-site sample set consisted of 8,981 samples of which 4,986 came from community controls and 3,995 from cases of severe or very severe pneumonia. The breakdown of *lytA* positivity by site and case-control status can be found in table 11.9. Overall, 7.0% of cases and 5.4% of controls were *lytA* positive, chi-squared p=0.002.

Table 11.9: *lytA* PCR positivity from whole blood samples among cases and controls by site in the Pneumonia Etiology Research for Child Health (PERCH) study

	Controls			Cases			
Site	N	*lytA* positive	(%)	N	*lytA* positive	(%)	Total
Kenya	751	48	6%	566	31	5%	1,317
The Gambia	608	47	8%	602	61	10%	1,210
Mali	715	38	5%	669	77	12%	1,384
Zambia	602	26	4%	524	32	6%	1,126
South Africa	963	98	10%	917	70	8%	1,880
Thailand	622	5	1%	224	2	1%	846
Bangladesh	725	6	1%	493	5	1%	1,218
Total	4,986	268	5%	3,995	278	7%	8,981

11.3.2.2 Characteristics of cases and controls in PERCH

Characteristics of cases and controls in the multi-site PERCH sample set can be found in table 11.10.

Table 11.10 Characteristics of cases and controls in the Pneumonia Etiology Research for Child Health (PERCH) multi-site sample set

Characteristic	Controls	(%)	Cases	(%)	p-value
Site:					
Kenya	751	15.1	566	14.2	
The Gambia	608	12.2	602	15.1	
Mali	715	14.3	669	16.8	
Zambia	602	12.1	524	13.1	
South Africa	963	19.3	917	23.0	
Thailand	622	12.5	224	5.6	
Bangladesh	725	14.5	493	12.3	<0.001
Female gender	2,491	50.0	1,704	42.7	<0.001
Age:					
2-5 months	1,534	30.8	1,625	40.7	
6-15 months	1,625	32.6	1,314	32.9	
16-59 months	1,827	36.6	1,056	26.4	<0.001
Nasopharyngeal carriage of pneumcoccus:					
by PCR	3,617	76.7	2,675	72.0	<0.001
by culture	3,419	69.5	2,086	53.0	<0.001
PCV-10-type by culture*	930	30.8	565	30.1	0.595
by culture or PCR	4,025	82.3	2,963	77.1	<0.001
Blood culture:					
positive for pneumococcus	-		45	1.1	
positive for another pathogen	-		119	3.0	
negative	-		3,831	95.9	-
Any virus in nasopharynx by multiplex PCR	2,716	57.5	2,932	79.1	<0.001
Pneumococcal conjugate vaccine status	2,469	50.8	2,068	53.4	0.014
Antibiotics prior to blood sample collection	111	2.4	1,571	41.4	<0.001
HIV infection	211	5.1	253	6.9	<0.001
Specimen extracted within a week of collection	1,488	29.8	1,072	26.8	0.002

*10-valent pneumococcal conjugate vaccine-type

11.3.2.3 *lytA* qPCR quantitative data in PERCH

The $\log_{10}$-transformed distribution of positive *lytA* qPCR results by case-control status and for cases with blood-culture confirmed IPD, can be seen in figure 11.8.

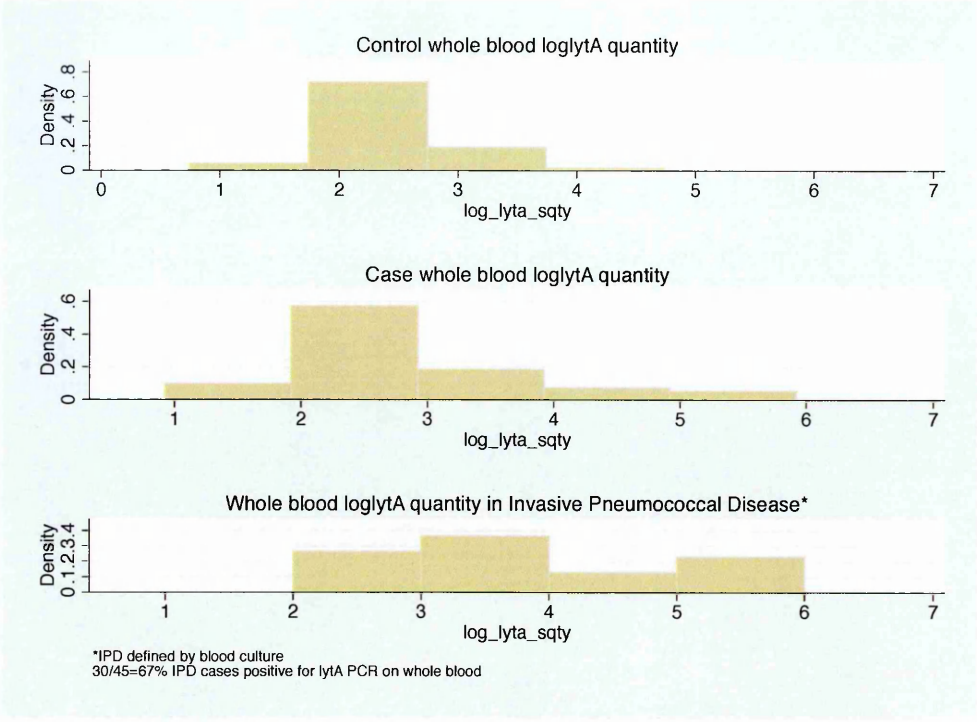


Figure 11.8: Distribution of log_{10}-transformed *lytA* concentration in log_{10} copies/mL, in whole blood samples in controls, in cases and in invasive pneumococcal disease (IPD) in the Pneumonia Etiology Research for Child Health (PERCH) multi-site sample set. IPD was defined by growth of *S. pneumoniae* from a blood culture. IPD cases are a subset of all cases.

The log_{10}-transformed concentration of *lytA* in whole blood for positive samples is displayed by participant characteristic in table 11.11. In order to further explore the potential contribution of nasopharyngeal carriage to blood *lytA* positivity we looked at the *lytA* concentration in blood by case-control and nasopharyngeal carriage status; these can be seen in figure 11.9. Among controls, the median *lytA* concentration in blood for those without demonstrated nasopharyngeal carriage of pneumococcus by culture or PCR was 2.07 log_{10} copies/mL (IQR 1.93-2.32 log_{10} copies/mL) and for those with carriage was

2.31 log_{10} copies/mL (IQR 2.07-2.67 log_{10} copies/mL), Wilcoxon rank-sum p=0.010. Among cases, the median *lytA* concentration in blood for those without nasopharyngeal carriage of pneumococcus was 2.31 log_{10} copies/mL (IQR 2.07-2.78 log_{10} copies/mL) and for those with carriage was 2.56 log_{10} copies/mL (IQR 2.18-3.23 log_{10} copies/mL), p=0.144.

Table 11.11: Distribution of pneumococcal concentration in positive whole blood samples by *lytA* qPCR in log_{10} copies/mL by participant characteristic in the Pneumonia Etiology Research for Child Health (PERCH) multi-site sample set.

Characteristic	Number *lytA* positive	Median log10 *lytA* quantity copies/mL	IQR		p-value*
Gender:					0.848
Male	265	2.39	2.10	2.95	
Female	281	2.42	2.08	2.81	
Age in tertiles:					0.196
2-5 months	188	2.37	2.10	2.71	
6-15 months	200	2.44	2.07	2.94	
16-59 months	158	2.45	2.09	3.14	
PCV-10 vaccination status§:					0.690
Not vaccinated	158	2.46	2.10	2.91	
Vaccinated	365	2.38	2.09	2.86	
Blood culture:					<0.001
No pneumococcus	248	2.47	2.14	3.01	
Pneumococcus	30	3.64	2.86	4.99	
Nasopharyngeal carriage by culture or PCR:					0.002
Negative	54	2.15	1.96	2.71	
Positive	477	2.43	2.11	2.91	
PCV-10 serotypes on nasopharyngeal culture§:					0.522
Absent	232	2.41	2.10	2.83	
Present	114	2.44	2.12	3.07	
Any respiratory virus by multiplex PCR:					0.247
Absent	176	2.37	2.05	2.81	
Present	338	2.43	2.11	2.86	
Antibiotics prior to collection of blood sample:					0.004
Absent	406	2.38	2.07	2.84	
Present	114	2.51	2.18	3.17	
HIV status:					0.015
Negative	420	2.36	2.10	2.85	
Positive	54	2.62	2.20	3.58	
Case-control status:					<0.001
Well community control	198	2.23	2.02	2.59	
Community control with URTI symptoms	70	2.37	2.06	2.81	
Hospitalised with severe pneumonia	165	2.46	2.13	2.99	
Hospitalised with very severe pneumonia	113	2.70	2.24	3.33	

Characteristic	Number *lytA* positive	Median log10 *lytA* quantity copies/mL	IQR		p-value*
Site:					<0.001
Kenya	79	2.34	2.10	2.71	
The Gambia	108	2.38	2.10	2.80	
Mali	115	2.77	2.31	3.47	
Zambia	58	2.33	2.12	2.61	
South Africa	168	2.28	2.07	2.72	
Thailand	7	1.95	1.65	2.93	
Bangladesh	11	1.97	1.75	2.10	
Sample collection period:					0.211
Jul-Dec 2011	60	2.38	2.14	2.60	
Jan-Jun 2012	152	2.48	2.15	3.00	
Jul-Dec 2012	115	2.40	2.04	2.96	
Jan-Jun 2013	144	2.36	2.11	2.76	
Jul-Dec 2013	66	2.32	2.07	3.04	
Jan-Jun 2014	9	2.07	1.83	2.52	
Days from specimen collection to DNA extraction:					0.569
Within 7 days	165	2.41	2.10	2.91	
More than 7 days	381	2.39	2.10	2.85	

*Wilcoxon rank-sum for binary categorical data, Kruskal-Wallis for >2 categories

§10-valent pneumococcal conjugate vaccine

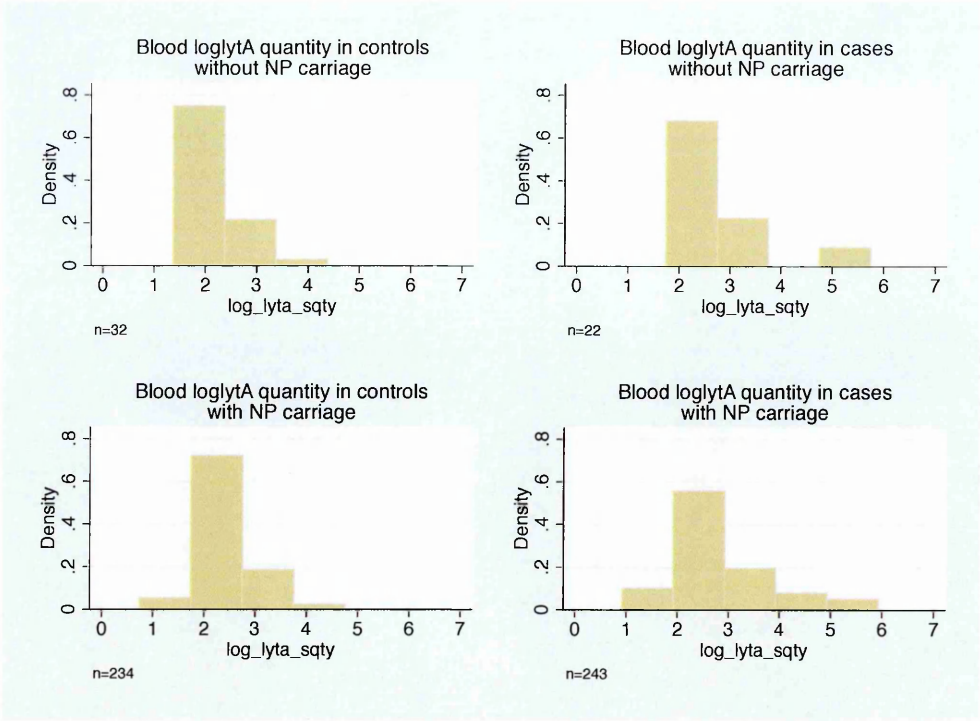


Figure 11.9: Whole blood *lytA* concentration in log_{10} copies/mL by case-control status and nasopharyngeal (NP) carriage status in the Pneumonia Etiology Research for Child Health (PERCH) sample set. Carriage was defined as culture of *S. pneumoniae* from a nasopharyngeal swab or *lytA* PCR positivity from a nasopharyngeal/oropharyngeal swab.

11.3.2.4 Correlation between *lytA* concentration in blood and nasopharyngeal compartments in PERCH

The positive correlation between *lytA* concentration in whole blood and *lytA* concentration in the nasopharynx, for children who were positive in both compartments, can be seen in figure 11.10.

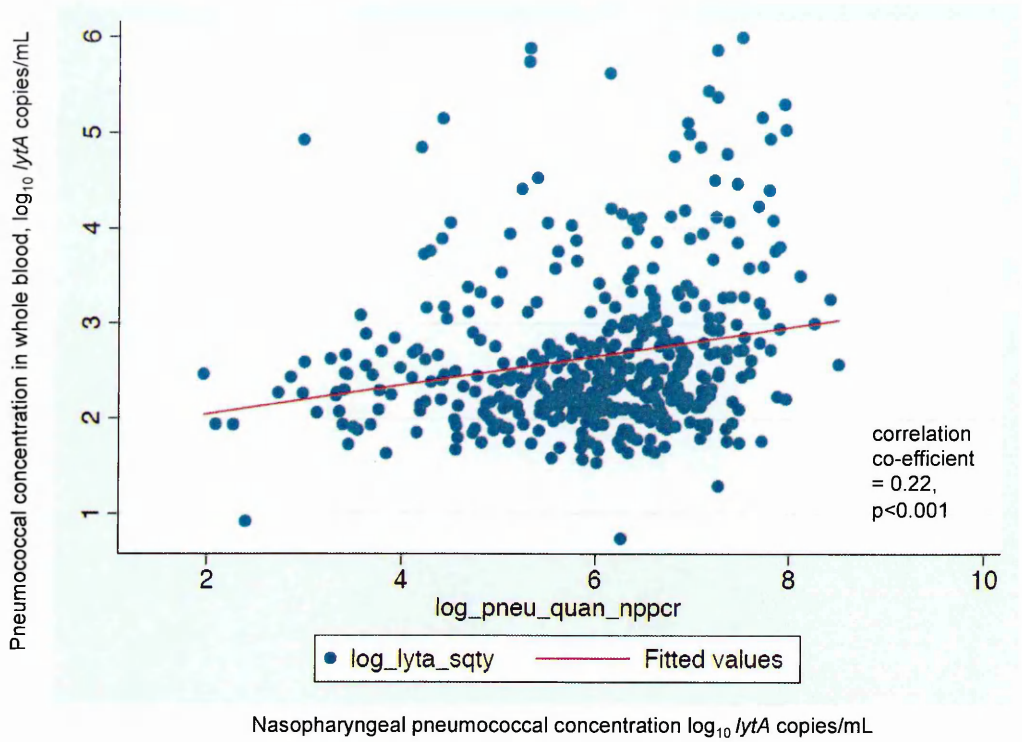


Figure 11.10: Correlation between pneumococcal concentration in the nasopharynx and in whole blood, for positive samples by *lytA* qPCR, in log_{10} *lytA* copies/mL, in the Pneumonia Etiology Research for Child Health (PERCH) multi-site sample set.

11.3.2.5 Exploring *lytA* qPCR assay performance in PERCH; categorising positive results

Of the thresholds suggested by the ROC curve from the Kilifi dataset, the threshold at ≥1000 copies/mL in quantity prior to adjustment for extraction:elution ratio, was selected to apply to the PERCH all-site dataset to categorise positive samples into strong and weak positives. This threshold had better specificity than the lower threshold and was closer to the limit of detection and linearity of the assay. Univariable and multivariable logistic regression was carried out to look for associations with strong positive and weak positive *lytA* PCR on whole blood. The results of the logistic regression for strong positive outcomes can be found in table 11.12 and for weak positive outcomes in table 11.13.

Table 11.12: Multivariable logistic regression for associations with a strong positive *lytA* PCR on whole blood in the Pneumonia Etiology Research for Child Health (PERCH) sample set. Variables associated with outcome in the univariable analysis with a p-value <0.1 and age, were included in the backwards stepwise multivariable model. Interaction terms significant at a p-value <0.05 by likelihood ratio test were included in the multivariable model.

Strong Positive Outcome *lytA* ≥1000 copies/mL		Univariable analysis					Multivariable analysis			
Variable	**n**	**OR**	**95% CI**		**p-value**	**Interaction terms**	**OR**	**95% CI**		**p-value**
Site:										
Zambia (reference)	8981	1.00				Zambia (reference)	1.00			
The Gambia		3.23	1.59	6.56	0.001		**3.05**	**1.46**	**6.37**	**0.003**
Mali		5.68	2.91	11.09	<0.001		**7.56**	**3.77**	**15.19**	**<0.001**
Kenya		1.72	0.80	3.69	0.163		1.65	0.73	3.75	0.233
South Africa		2.67	1.34	5.34	0.005		**2.86**	**1.44**	**5.65**	**0.003**
Thailand		0.26	0.06	1.21	0.086		0.45	0.11	1.83	0.264
Bangladesh		0.09	0.01	0.72	0.023		**0.16**	**0.03**	**0.87**	**0.034**
HIV	7854	2.98	1.89	4.70	<0.001	*Interaction between HIV and carriage*				
						Without HIV, without carriage (reference)	1.00			
						With HIV, without carriage	0.65	0.04	12.06	0.774
						Without HIV, with carriage (reference)	1.00			
						With HIV, with carriage	**3.24**	**1.96**	**5.34**	**<0.001**
						With vs without carriage, without HIV	1.79	0.72	4.49	0.213
Case status	8981	2.32	1.70	3.16	<0.001		**2.15**	**1.50**	**3.09**	**<0.001**
PCV vaccination status*	8734	2.19	1.57	3.05	<0.001					
Antibiotic prior to sample collection	8436	1.38	0.98	1.95	0.069					
Nasopharyngeal carriage	8734	2.84	1.64	4.92	<0.001					
Age in tertiles:	8981					*Interaction between age and carriage*				
1-5 months of age (reference)		1.00				1-5 months of age without carriage (reference)	1.00			
6-15 months of age		1.36	0.94	1.97	0.107	6-15 months of age without carriage	0.16	0.01	2.87	0.211
16-59 months of age		1.36	0.94	1.98	0.105	16-59 months of age without carriage	2.00	0.53	7.56	0.308
						1-5 months of age with carriage (reference)	1.00			
						6-15 months of age with carriage	**1.71**	**1.10**	**2.66**	**0.017**
						16-59 months of age with carriage	**2.02**	**1.29**	**3.18**	**0.002**
						With vs without carriage, age 1-5 months	1.79	0.72	4.49	0.213
Female gender	8980	1.04	0.78	1.40	0.779					
Any respiratory virus in nasopharynx	8430	1.04	0.75	1.45	0.808					
Days from specimen collection to DNA	8278	1.00	0.99	1.00	0.226					
DNA extraction period	8822	0.91	0.80	1.03	0.120					

*PCV = pneumococcal conjugate vaccine

Table 11.13: Multivariable logistic regression for associations with a weak positive *lytA* PCR on whole blood in the Pneumonia Etiology Research for Child Health (PERCH) sample set. Variables associated with outcome in the univariable analysis with a p-value <0.1 and case-control status, were included in the backwards stepwise multivariable model. Interaction terms significant at a p-value <0.05 by likelihood ratio test were included in the multivariable model.

Weak Positive Outcome *lytA* <1000 copies/mL	Univariable analysis						Multivariable analysis			
Variable	**n**	**OR**	**95% CI**		**p-value**	**Interaction terms**	**OR**	**95% CI**		**p-value**
PCV vaccination status*	8564	2.23	1.77	2.80	<0.001					
Nasopharyngeal carriage	8563	2.08	1.49	2.90	<0.001	*Interaction between carriage and case-control status*				
						No carriage, in controls (reference)	1.00			
						Carriage in cases	**2.87**	**1.64**	**5.05**	**<0.001**
						Carriage in controls	1.49	0.97	2.31	0.071
HIV	7704	1.80	1.23	2.64	0.003					
Site:	8803									
Zambia (reference)		1.00					1.00			
The Gambia		1.49	1.03	2.17	0.035		1.42	0.97	2.07	0.070
Mali		0.84	0.56	1.27	0.408		0.90	0.60	1.36	0.622
Kenya		1.06	0.72	1.57	0.768		1.06	0.71	1.58	0.768
South Africa		1.61	1.14	2.27	0.006		**1.67**	**1.18**	**2.36**	**0.004**
Thailand		0.13	0.05	0.33	<0.001		**0.15**	**0.06**	**0.37**	**<0.001**
Bangladesh		0.18	0.09	0.37	<0.001		**0.17**	**0.08**	**0.35**	**<0.001**
Antibiotic prior to sample collection	8263	1.02	0.78	1.34	0.867					
Case status	8803	1.01	0.82	1.25	0.932	**Case status, no carriage**	**0.46**	**0.24**	**0.88**	**0.019**
Age in tertiles:										
1-5 months of age (reference)										
6-15 months of age	8803	1.08	0.84	1.37	0.550		1.15	0.89	1.47	0.282
16-59 months of age		0.76	0.58	0.99	0.040		0.89	0.67	1.17	0.396
DNA extraction period	8645	0.98	0.90	1.07	0.634					
Female gender	8802	1.10	0.89	1.35	0.384					
Days from specimen collection to DNA extraction	8104	1.00	1.00	1.00	0.404					
Any respiratory virus in nasopharynx	8265	0.90	0.72	1.13	0.367					

*PCV = pneumococcal conjugate vaccine

11.3.2.6 Test parameters by site, in PERCH

The sensitivity and specificity of strong positive outcome at ≥1000 copies/mL was calculated for each PERCH site and tabulated in table 11.14. The sensitivities and specificities for African sites can also be seen in figure 11.11, with 95% confidence intervals. The positive and negative predictive values for the sensitivity and specificity at each site can be seen in table 11.15. A range of possible prevalence for IPD is displayed, including the prevalence of positive *lytA* results among cases at a cut-off of ≥1000 copies/mL for each African site. Calculation of test parameters for Asian sites was limited to specificity as there were no cases of IPD as defined by blood culture in Thailand or Bangladesh.

Table 11.14: Test parameters by site in the Pneumonia Etiology Research for Child Health (PERCH) sample set. Sensitivity and specificity of a cut-off value of 1000 copies/mL, before adjustment for extraction:elution ratio, for a positive *lytA* qPCR on whole blood. Sensitivity is set against invasive pneumococcal disease (IPD) by blood culture and specificity is among controls.

Strong positives as test positives with a cut-off of ≥1000 copies/mL in raw quantity from *lytA* qPCR										
Site	**Test positives**	**IPD§**	**Sensitivity**	**95% CI**		**Test negatives**	**Controls**	**Specificity**	**95% CI**	
Kenya	3	4	75%	19%	99%	739	751	98.4%	97.2%	99.2%
The Gambia	4	10	40%	12%	74%	596	608	98.0%	96.6%	99.0%
Mali	17	23	74%	52%	90%	698	715	97.6%	96.2%	98.6%
Zambia	4	7	57%	18%	90%	601	602	99.8%	99.1%	99.996%
S. Africa	0	1	0%	0%	*98%	944	963	98.0%	96.9%	98.8%
Thailand	0	0	-	-	-	620	622	99.7%	98.8%	99.96%
Bangladesh	0	0	-	-	-	725	725	100.0%	*99.5%	100.0%
p-value Fisher's exact test			0.224					<0.001		
p-value Fisher's exact test for heterogeneity, African sites only								0.004		

*one-sided 97.5% confidence interval; § Invasive Pneumococcal Disease

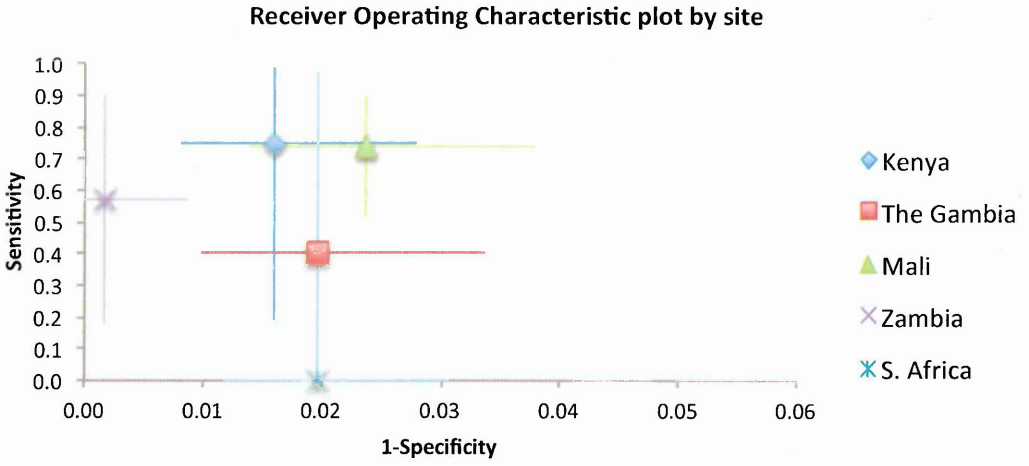


Figure 11.11: Sensitivity and specificity of a cut-off of 1000 copies/mL in *lytA* qPCR, by site. Sensitivity and specificity of a cut-off of 1000 copies/mL, before adjustment for DNA extraction:elution ratio, for a positive *lytA* qPCR on whole blood, with 95% confidence intervals, for African sites in the Pneumonia Etiology Research for Child Health (PERCH) sample set.

Table 11.15: Test parameters of *lytA* qPCR on blood for African sites at a cut-off of ≥1000 copies/mL. Positive and negative predictive values (PPV, NPV) for *lytA* qPCR on blood in the Pneumonia Etiology Research for Child Health (PERCH) sample set using a cut-off of ≥1000 copies/mL before adjustment for the DNA extraction:elution ratio, for a range of prevalence in invasive pneumococcal disease (IPD). Actual prevalence of *lytA* PCR positivity at this cut-off among cases for each site is shown in red. Sensitivity is set against IPD by blood culture and specificity is among controls.

Test parameters: cut-off ≥1000 copies/mL				
Kenya				
Prevalence (%)	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**
0.1%	75%	98.4%	4.48%	99.97%
0.5%	75%	98.4%	19.06%	99.87%
1.0%	75%	98.4%	32.13%	99.74%
1.4%	75%	98.4%	39.96%	99.64%
5.0%	75%	98.4%	71.16%	98.68%
10.0%	75%	98.4%	83.89%	97.25%
The Gambia				
Prevalence (%)	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**
0.1%	40%	98.0%	1.96%	99.94%
0.5%	40%	98.0%	9.13%	99.69%
1.0%	40%	98.0%	16.81%	99.39%
3.7%	40%	98.0%	43.45%	97.70%
5.0%	40%	98.0%	51.28%	96.88%
10.0%	40%	98.0%	68.97%	93.63%
Mali				
Prevalence (%)	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**
0.1%	74%	97.6%	2.99%	99.97%
0.5%	74%	97.6%	13.42%	99.87%
1.0%	74%	97.6%	23.75%	99.73%
5.0%	74%	97.6%	61.87%	98.62%
7.5%	74%	97.6%	71.43%	97.89%
10.0%	74%	97.6%	77.41%	97.13%
Zambia				
Prevalence (%)	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**
0.1%	57%	99.8%	22.20%	99.96%
0.5%	57%	99.8%	58.88%	99.78%
1.0%	57%	99.8%	74.22%	99.57%
1.7%	57%	99.8%	83.13%	99.26%
5.0%	57%	99.8%	93.75%	97.78%
10.0%	57%	99.8%	96.94%	95.43%
South Africa				
Prevalence (%)	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**
0.1%	0%	98.0%	0.00%	99.90%
0.5%	0%	98.0%	0.00%	99.49%
1.0%	0%	98.0%	0.00%	98.98%
2.7%	0%	98.0%	0.00%	97.25%
5.0%	0%	98.0%	0.00%	94.90%
10.0%	0%	98.0%	0.00%	89.82%

11.3.2.7 Effect on PERCH study outcomes

Taking a cut-off of ≥1000 copies/mL in *lytA* qPCR on blood, would give an additional 87 diagnoses of IPD on top of 28 diagnoses by blood culture and *lytA* PCR and 17 diagnoses by blood culture alone for the PERCH study, *ie*. the PERCH study would go from 45 cases of IPD to 132 cases, in this dataset.

11.4 Discussion

The *lytA* qPCR assay was positive among 1-10% of controls in seven different countries in Africa and Asia, from which we can conclude that the assay is not 100% specific for the diagnosis of IPD as has been previously reported.[15,148] Why is this? Both the Italian group and the South African group used the same *lytA* primers and probe as those used in the Kilifi and PERCH datasets, published by the CDC and in widespread use globally.[125] Even the extraction method used for the Kilifi dataset, the manual Qiagen method, was also used by both Rouphael *et al* and Azzari *et al*. The sample type used in Italy was whole blood but in South Africa it was serum. Given the similarity in the laboratory methods between Azzari *et al* and the data presented here, it is likely that differences lie in the populations tested.

The Italian group tested 147 healthy children aged 6 months to 16 years, with a median age of 4.9 years, who were attending hospital for allergies or celiac disease. The one-sided 97.5% confidence interval around a point estimate of 100% among 147 samples is 97.52%-100%. The South African group tested 100 HIV-negative children with a mean age of 5.8 years who were being followed up in a study of PCV-9. The one-sided 97.5% confidence interval around a point estimate of 100% among 100 samples is 96.37%-100%. In both the Kilifi dataset and in the multi-site PERCH study, the median age of controls

was 1.1 years, with an IQR of 6 months to 2.1 years, so the population we studied was younger than the previously published data.

Half of the children studied in Italy and in South Africa previously were found to be carrying pneumococcus in the nasopharynx, by PCR in Italy and by culture in South Africa. In the PERCH study 82% of controls were found to be carrying pneumococcus by culture or PCR (70% by culture alone; 77% by PCR alone), in the Kilifi dataset 89% of controls were carrying pneumococcus by culture or PCR (78% by culture alone; 81% by PCR alone). So our data come from a younger population with a higher prevalence of nasopharyngeal carriage of pneumococcus.

Among PERCH controls who were PCR-positive for pneumococcus in the nasopharynx, the median nasopharyngeal concentration was 5.77 $\log_{10}$ copies/mL (IQR 4.96-6.43 $\log_{10}$ copies/mL) whereas in the Italian study the distribution was reported as a mean value of 2.4 log copies/mL with an IQR of 2.2-3.7 log copies/mL. Even allowing for differences in the standards used for qPCR, it is likely that the nasopharyngeal concentration of pneumococcus was higher in our population.

There are likely to be other differences between the Italian population and the PERCH study populations. The PERCH project was carried out in resource-poor countries among populations with low socio-economic opportunity. It is probable that the Italian children who were attending hospital for allergies or celiac disease came from a higher socio-economic background.

The difference between the PERCH results and the previous South African study cannot only be a difference in populations, as South Africa was one of the sites in the PERCH project and had positive *lytA* qPCR results among 10% of controls in PERCH. Both

populations came from Soweto, in Johannesburg, but the previously studied population were older and were selected on the basis of their caregivers responding to an invitation to come for a booster pneumococcal conjugate vaccine. The previously published South African study used serum for the *lytA* PCR, rather than whole blood, and this might also contribute to the difference in results. Serum does not contain white blood cells, which could be where pneumococcal DNA is concentrated.

The *lytA* qPCR method was published by the CDC in 2007 [125] and is in widespread use globally, so it is possible that other investigators have also used the assay in control groups and not published the results because they found positive results in controls. An assumption may have been made that there was lab error, which could have lead to publication bias. The Kilifi sample set and the PERCH project between them used seven different laboratories and found positives among every control population.

We set out to examine why the *lytA* qPCR assay is not specific. Only a strong positive outcome was associated with case status. Strong positives were also associated with HIV positivity in the presence of nasopharyngeal carriage of pneumococcus, which is biologically plausible because it is known that HIV is a strong risk factor for IPD.[43,74] If we assume for a moment that strong positives are true positives, then what are the weak positives? Some of these could be true weak positives in cases of IPD, perhaps partly treated IPD. Some could be false positives in terms of the diagnosis of IPD, but what are they detecting? Does pneumococcal "DNAemia" occur among healthy children as has been suggested in the past? [106] It has been thought that the false positives in the Dagan study may have been due to the non-specificity of the *ply* target gene for pneumococcus. However, non-pneumococcal streptococci would not be expected to be accessing the bloodstream sufficiently to cause positive PCR assays on blood samples either, and if they

do so then we should expect pneumococci to do the same. Analytical specificity aside, it is likely that Dagan *et al*'s study was detecting pneumococcal DNA in the bloodstream of at least some of the healthy controls in whom *ply* PCR was positive.

Of course we must consider whether the technology of the methods could have introduced contamination resulting in false positives. This has been suggested [15] as a potential problem particularly of nested PCR,[106,211] where amplicons are accessed during the assay unlike the real-time PCR assay used for PERCH, but remains a possibility with any amplification method. To investigate this we looked at sample positivity by sample collection date, DNA extraction date, PCR date and laboratory technologist in the Kilifi sample set, and included DNA extraction period in the univariable and multivariable analyses for both the Kilifi and PERCH datasets. In Kilifi there was an association with date and sample positivity, and given the expected collinearity between date data, extraction period was selected for inclusion in the logistic regression models because positivity varied most significantly by extraction period. There was variability in positivity between laboratory technologists but this was explained by differences over time.

It is possible that the association with extraction period represents the laboratory improving in molecular methods with time. Certainly in Kilifi we made improvements in laboratory quality assurance over the time period of the project, with implementation of a Standard Operating Procedure aimed at reducing the potential for intra-laboratory contamination (for example; unilateral workflow in the molecular laboratory, cleaning of laboratory workspace and equipment with commercially-available DNA-degrading solutions, frequent changing of gloves during molecular procedures, spinning down the contents of sample vials prior to opening them). However, weak positive samples and positive control samples were a feature of all seven laboratories testing samples in the

PERCH project and it seems very unlikely that false positives due to intra-laboratory contamination could explain all of these data.

Could there be another explanation for changes in positivity over the time course of the project? PCV-10 was introduced to the Kenyan National Immunisation schedule in January of 2011, shortly after data collection began for the larger Kilifi dataset (November 2010) and shortly before the Kilifi arm of the PERCH project began (August 2011). Increasing use of PCV-10 in the community led to decreasing likelihood of IPD, which could account for decreasing numbers of positive samples with time, but it is hard to see how the vaccine might have decreased weak positive *lytA* qPCR results over time but not strong positives.

Nasopharyngeal carriage of pneumococcus is known to be a pre-requisite for IPD [66,67] and it has been the bane of IPD diagnostics for decades, especially among children where diagnosis is particularly difficult. It has therefore been assumed that nasopharyngeal carriage may cause false positive PCR assays when testing blood in the diagnosis of IPD and previous investigators [15,148] were careful to include children with demonstrated nasopharyngeal carriage of pneumococcus when examining the specificity of *lytA* PCR for IPD. Our multivariable logistic regression model found significant interactions between carriage and HIV, and carriage and age, in strongly positive *lytA* qPCR. These are biologically plausible associations that are reassuring. When weak positive *lytA* qPCR was the outcome, we found significant associations between carriage and outcome for cases only, not among controls. This implies that carriage is associated with true positives but fails to make a case for either nasopharyngeal carriage or lab error causing false positives in the diagnosis of IPD. Biologically, nasopharyngeal pneumococcal carriage must be the source for pneumococcal "DNAemia" if this occurs.

If we suppose that positive *lytA* PCR results among healthy controls are due to true pneumococcal "DNAemia", then we have another question to consider; what is invasion and what is invasive disease? Do pneumococci invade through the mucosa of the nasopharynx regularly enough to explain our findings among controls? Are we detecting organisms within white blood cells, already taken care of by the immune system? In that case, we further suppose that IPD occurs only when immune escape occurs after the relatively common invasion event is (rarely) unable to be contained and pneumococci are able to divide and grow in the bloodstream, allowing blood cultures to become positive.

A noteworthy finding from these data is the difference in *lytA* positivity by region. There is a clear difference in positivity between the African and the Asian sites within the PERCH project. There were no cases of IPD as defined by blood culture positivity in either of the two sites in Thailand or the two sites in Bangladesh that contributed to the PERCH project, and there were very few *lytA* positive samples from these sites, with just 1% of cases and 1% of controls testing positive from both Thailand and Bangladesh. We infer from this that there must be genetic differences in susceptibility to IPD between African and Asian populations. Nevertheless, nasopharyngeal carriage was common in Thailand (66% of controls and 57% of cases) and Bangladesh (87% of controls and 74% of cases), so the difference must be in either the likelihood of an invasion event or the handling of an invasion event, or both. *lytA* positivity was significantly lower among the Asian sites than African sites but the same among Asian controls as Asian cases, so if "DNAemia" is a real phenomenon then it must be occurring in Asia too, but at a lower frequency.

A clear limitation of this assay is the lower limit of reliable detection is approximately 1000 copies/mL, or 100 copies/reaction, in raw concentration, equating to 500 copies/mL of whole blood after allowing for the extraction:elution ratio. Because the median *lytA*

concentration in blood was below this, at 178 copies/mL among controls and 338 copies/mL among cases, we expect that there would have been more positives if the assay were more sensitive. It also means that the concentrations below 500 copies/mL are subject to lack of precision due to being below the lower limit of linearity of the quantitative PCR. The fact that we see biologically plausible associations using thresholds to categorise the data into strong and weak positives at 1000 copies/mL in raw concentration suggests that there is still useful information contained in the data despite these limitations. Given the low median concentrations of *lytA* in blood, it would not have been feasible to set the threshold higher.

The *lytA* qPCR assay in blood is not 100% specific in the diagnosis of IPD in all populations and should not be used as a diagnostic assay for clinical care without careful examination of test parameters population by population. For the PERCH project, setting a threshold of 1000 copies/mL and taking all positives above this threshold to be indicative of IPD would increase the number of cases of IPD from 45 by blood culture to 132 by blood culture or *lytA* qPCR. This could be a very helpful augmentation of endpoints for the PERCH project overall in assigning causality of pneumonia. It is possible that the *lytA* qPCR assay detects pneumococcal "DNAemia", or pneumococci that are no longer viable after being lysed or phagocytosed by the immune system. More work is needed to examine what happens when pneumococci breach the nasopharyngeal mucosa, how often this occurs, and the genetic differences in handling such invasion events between different human populations. Although it is important to understand the pathophysiology, the fact remains that pneumococcal PCR on blood specimens is non-specific in the diagnosis of IPD in children.

Chapter Twelve

12 Discussion of contributions to the field

Nasopharyngeal carriage of pneumococcus appears to be a nearly ubiquitous part of being human.[23] Carriage is a required antecedent to pneumococcal pneumonia and invasive pneumococcal disease (IPD), both of which are fortunately much less common occurrences.[67,202] Young children are the main reservoir from which transmission occurs to others. Young children are also the group most at risk for IPD and pneumococcal pneumonia and yet they are the group among whom these diseases are the most difficult to diagnose.

Nasopharyngeal pneumococcal concentration does not hold promise as a diagnostic assay for pneumonia in children as it apparently does in adults,[126] at least not in our population. Not only was there considerable overlap in nasopharyngeal pneumococcal concentration between well community controls, community controls with upper respiratory tract (URTI) symptoms and children admitted to hospital with severe or very severe pneumonia, but the community controls with URTI symptoms had statistically significantly greater pneumococcal concentration in the nasopharynx than did children with pneumonia. This was not just an antibiotic effect as it remained when the analysis was restricted to those in whom there was no evidence for prior antibiotic use, although there was probably still some confounding by antibiotic use remaining. These results differ from those previously published in Vietnam [99] where all swabs were collected prior to antibiotic administration.

It is worth noting that the previously published work was carried out in South-East Asia, where pneumococcus seems to be handled differently, presumably due to host genetic

differences. In the multi-site PERCH study, children at Asian sites were less likely to carry pneumococcus than children at African sites, although carriage was still common. More remarkable though, was the fact that IPD, as diagnosed by blood cultures, was so rare that no cases were detected among children with severe or very severe pneumonia at the Asian sites, and *lytA* PCR on whole blood was dramatically less likely to be positive than in Africa.

Respiratory viruses are also a ubiquitous part of the human experience. As discussed in section 2.1.3 of the introduction, ecological studies,[71,73] basic science experiments on human respiratory epithelial cell lines [77,83] and animal models [82,89,95] all tell us that bacteria in the nasopharynx are affected by the presence of respiratory viruses. Clinically, we note that severe bacterial pneumonia may follow respiratory virus infection.[75] We can hypothesise that respiratory virus infection in humans increases the nasopharyngeal pneumococcal concentration and that the concentration of pneumococcus in the nasopharynx is associated with the likelihood of progression to pneumonia or IPD, as well as the likelihood of transmission of carriage.

Our study is the first longitudinal study of nasopharyngeal pneumococcal concentration by qPCR in children with episodes of respiratory virus infection. The concentration of pneumococcus in the nasopharynx of children <5 years of age increased 4-fold with the onset of a symptomatic respiratory virus infection with RSV or rhinovirus. The rise in pneumococcal concentration associated with respiratory viral infection was surprisingly modest but may be important for the risk of developing IPD or for transmission of pneumococcal carriage at a population level. The rise in pneumococcal concentration associated with respiratory viral infection may also be dependent on the pneumococcal serotype or strain; some pneumococci may be more likely to respond to respiratory viral

infection with an increase in concentration and in expression of pathogenicity than others. In our study there was a less marked decrease in pneumococcal concentration after the episode of viral infection that did not reach statistical significance. This may be an issue of timing; we took the mean pneumococcal concentration during the episode of viral infection relative to the mean pneumococcal concentration after the episode of viral infection. If the pneumococcal concentration remains elevated subsequent to viral clearance then it will be necessary to measure the time it takes for pneumococcal concentration to return to baseline in the future.

We don't know what the relationship between nasopharyngeal carriage density and IPD/pneumonia or between carriage density and transmission really is. If we assume that there is a sigmoid relationship, then we can postulate a threshold effect (as in figure 12.1). For a 4-fold increase in pneumococcal concentration in response to respiratory viral infection, such as we found in our longitudinal study, to have an important population-level effect, this increase would need to meet a threshold for invasion or transmission. However, if the steep part of the sigmoid curve occurs in the centre of the population distribution of carriage density, then the relationship may be effectively linear – or the relationship might not be sigmoid at all. If the threshold for transmission or for invasion occurs near the mean of the population normal curve for carriage density, then it would be possible to examine this. An animal model could be used to examine the relationship between carriage density and IPD, but the danger is that it might not hold true for humans. The relationship with transmission could be examined in a human longitudinal study but would need large enough numbers, especially if pneumococcal serotype was to be taken into consideration.

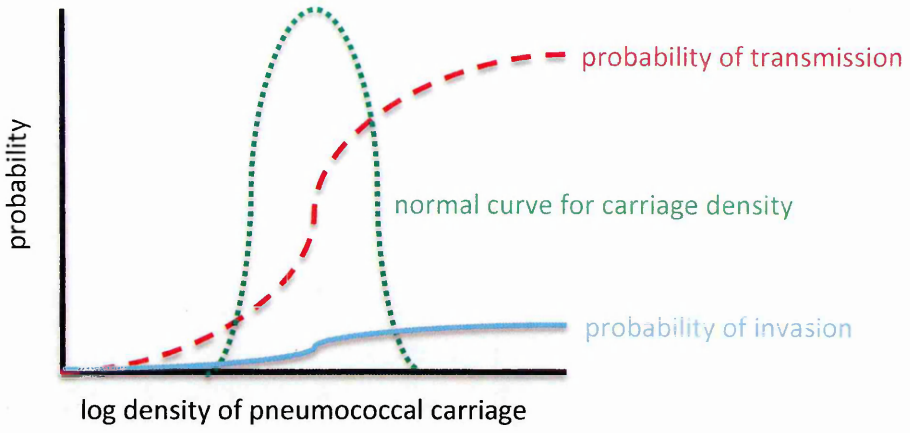


Figure 12.1 Sketch of possible relationships between nasopharyngeal pneumococcal concentration and transmission or invasion

Quantitative PCR may help us to understand the relationship between carriage density and transmission or invasion. However, if the threshold for transmission or invasion lies close to the lower limit of detection of the qPCR assay then quantitative results are difficult to use because the assay effectively becomes binary; detection or no detection. Complicating this is the fact that qPCR and culture appear to be complementary in detecting and quantifying nasopharyngeal carriage; qPCR is more sensitive, but we have found in this thesis, as have others,[175] some samples may be culture positive and PCR negative as well as vice versa. Presumably this is due to the ability of PCR to detect non-viable organisms, and to sample handling and PCR inhibition. An additional complication is that qPCR assays are not well standardised across laboratories, so a threshold studied in one laboratory does not necessarily correspond to a threshold studied in another laboratory unless efforts are made to calibrate standards across laboratories, as for the PERCH project.

The pattern of colonisation of the nasopharynx by pneumococcus after vaccination with PCV-10 is complex. The graphs of nasopharyngeal pneumococcal concentration by serotype among individual children over six months post PCV-10 demonstrate the striking changes in pneumococcal concentration that occur within individuals and the dominance of the prevailing serotype. Such changes in concentration could impact the risk of transmission or invasion.

The strength of the pneumococcal microarray is that it allows us to see these complex patterns of multiple serotype carriage, but there are two important weaknesses in the method. One is the artefact or step in the data that occurs with culture negative samples. Examination of culture-positive samples alone will introduce bias if culture-negative, *lytA*-positive samples are different in the quantity and pattern of pneumococcal serotypes present. The second is that the preliminary culture step introduces a bias for strains that multiply more effectively than other strains during the culture process. Quantitation of serotypes using a combination of *lytA* qPCR and pneumococcal microarray yields quantitative data that are complicated by the presence of many zero quantities, meaning data analysis is not straightforward.

Our microarray project found that VT pneumococcus did not substantially change post PCV-10, while NVT pneumococcus and total pneumococcal concentration increased, a somewhat weaker finding than we had expected to see. Perhaps a more important contribution to the field is the advance in understanding the complexity of the data generated by the approach. The pneumococcal microarray has been advocated by the PneuCarriage project [171] as the method of choice for quantitative data that detects multiple serotype carriage and the next generation of vaccine studies with carriage outcomes are lining up to use it. Use of the pneumococcal microarray for cross-sectional

surveys rather than for longitudinal studies would be more straightforward but longitudinal studies are epidemiologically stronger for studying vaccine impact. We need to continue to innovate to find ways to study the nasopharyngeal environment more accurately.

The use of *alu* qPCR to adjust nasopharyngeal swab specimens for the amount of human specimen present in a vial of transport medium was a novel approach. Using this assay enabled us to detect two specimens that did not contain any human sample and exclude them from the analysis in the longitudinal study of pneumococcal concentration before, during and after respiratory virus infection. In this study, the analysis without *alu* adjustment yielded results with a similar order of magnitude to the adjusted analysis, but a slightly stronger effect; a 7-fold rather than a 4-fold rise in pneumococcal concentration. In the study of nasopharyngeal pneumococcal concentration among healthy community controls, community controls with URTI symptoms and cases of severe or very severe WHO-defined pneumonia, the analysis without adjustment for *alu* concentration again did not remarkably change the results or the conclusions drawn.

For examination of pneumococcal concentration before and after administration of the pneumococcal conjugate vaccine, use of *alu* qPCR to adjust the pneumococcal concentration for the concentration of human specimen present made only small changes to the results of the analysis of VT, NVT and total pneumococcal concentration over the six-month period post vaccination. VT pneumococcal concentration did not change over time whether or not *alu* adjustment was made; NVT pneumococcal concentration increased significantly whether or not *alu* adjustment was made. Total pneumococcal concentration did not change over time when unadjusted for *alu* concentration but did appear to increase over time post vaccination when *alu* adjustment

was made. In the study of changes in nasopharyngeal pneumococcal concentration in response to vaccine, rayon nasopharyngeal swabs had been used and in the other two studies nasopharyngeal flocked swabs were used, with which it may be easier to collect a standard sample volume.

Considering these three studies together, we can most probably conclude that if adequate training and monitoring of swab collectors is performed, a standard volume of transport medium used and ideally if nasopharyngeal flocked swabs are used, then adjustment for the concentration of human DNA in the swab transport medium is not necessary. An alternate conclusion is that *alu* qPCR was not an adequate method for adjusting the pneumococcal concentration for the quantity of human specimen present in swab transport medium. Considering that our validation study showed that human DNA concentration, as measured by *alu* qPCR, increased with improved swab collection and handling in volunteers; that human DNA concentration correlated with protein concentration; and that in the studies discussed above the direction of the effect was the same with and without adjustment for human DNA, we think that the method was reasonable.

It is noteworthy that using a threshold for positivity of 40 cycles, the *lytA* qPCR was positive on most nasopharyngeal swab samples in STGG studied, surpassing the proportion positive by culture or by PCR from swabs in viral transport medium. In chapter eight of this work, we have described data from the *lytA* qPCR performed on a subset of nasopharyngeal swabs in STGG collected in the PERCH project in Kilifi. The prevalence of carriage among these swabs in STGG was 98.6% but among nasopharyngeal swabs collected into Copan UTM and assayed using a 33-plex PCR from Fast-Track Diagnostics (which includes a *lytA* target) was 80.8% among the same individual study participants.

Even among swabs collected in viral transport medium, in the study of pneumococcal concentration before, during and after acquisition of respiratory viral infections, swab positivity was 93.4%. Detection of nasopharyngeal carriage by PCR rather than by the previous WHO standard culture is becoming more common and may lead to a paradigm shift in how we view nasopharyngeal carriage. Perhaps in some populations like ours, pneumococcal carriage is normally so common that with sensitive detection methods prevalence approaches 100%. Deciding at what amplification cycle threshold or quantitative value to set the cut-off for a positive PCR assay impacts upon the estimation of carriage prevalence by PCR.

If the CDC-*lytA* PCR primer and probe set that we used was not 100% specific for pneumococcus, this could explain high carriage prevalence detected by PCR compared to culture. As detailed in chapter two, various groups have found 100% analytical specificity for pneumococcus with a *lytA* PCR among 82 *S. pseudopneumoniae* or other closely related non-pneumococcal streptococci.[125,136,138,139] Testing the assay on a well-characterised collection of *S. pseudopneumoniae*, *S. mitis* and *S. oralis* known to contain the autolysin gene from around the world would be more reassuring but such a collection does not, to our knowledge, exist. Assuming pneumococcal carriage in our population of young children is nearly universal by PCR, we should consider whether associations with carriage hold true, or whether associations with carriage are only true for carriage at high enough levels to be culture-positive. Carriage appears to be more of a continuum than a binary phenomenon in our population.

An important contribution to the study of pneumococcal disease is the finding that *lytA* qPCR on whole blood samples is not specific for the diagnosis of invasive pneumococcal disease among children under 5 years of age. A cut-off can be set to improve specificity so

that the assay might be helpful in the PERCH study model, but in order to exclude positive results among all controls all useful sensitivity is lost. Disappointingly, the assay is therefore not of clinical utility in the diagnosis of IPD among children in low and lower-middle income countries, the group in whom we most need improved diagnostics.

In 1998, Dagan et al reported 17% of 202 healthy controls were positive for *ply* PCR from serum samples; children were significantly more likely to be positive than adults, and children in whom nasopharyngeal carriage was detected by culture were significantly more likely to be positive than those who were not carriers. The authors postulated that pneumococcal DNA was present from direct bloodstream invasion or by phagocytosis of organisms by lymphoid cells that then entered the bloodstream. They pointed out that the development of systemic immune responses to pneumococcal types after nasopharyngeal carriage lends weight to this hypothesis.[106] Subsequently, *ply* has been found in non-pneumococcal streptococci, and the non-specificity of the assay for IPD has been largely overlooked because of the non-specificity for pneumococcus itself.

However, in light of our findings of the non-specificity of the *lytA* assay for IPD across seven different populations and in seven different laboratories, it appears that healthy children in developing countries frequently have pneumococcal DNA circulating in their bloodstream, at levels not dissimilar to those seen in children with disease. Further work is needed to determine how this occurs, how often this occurs, and the relationship between pneumococcal invasion resulting in harmless "DNAemia" and that resulting in IPD with serious morbidity and mortality. Although these questions are of considerable biological interest, answering them is unlikely to change the fact that among children, PCR on blood specimens is non-specific in the diagnosis of IPD.

Chapter Thirteen

13 References

1. Liu L, Johnson HL, Cousens S, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 2012;379:2151-61.

2. Scott JA, Brooks WA, Peiris JS, Holtzman D, Mulholland EK. Pneumonia research to reduce childhood mortality in the developing world. J Clin Invest 2008;118:1291-300.

3. Scott JAG, Hammitt LL, Bwanaali T, et al. Impact of introducing 10-valent pneumococcal conjugate vaccine in Kenya on invasive pneumococcal disease among children aged <5 years. 8th International Symposium on Pneumococci and Pneumococcal Diseases; 2012 March 11–15, 2012; Iguaçu Falls, Brazil.

4. Levine OS, O'Brien KL, Deloria-Knoll M, et al. The Pneumonia Etiology Research for Child Health Project: a 21st century childhood pneumonia etiology study. Clin Infect Dis 2012;54 Suppl 2:S93-101.

5. Hammitt LL, Kazungu S, Morpeth SC, et al. A preliminary study of pneumonia etiology among hospitalized children in Kenya. Clin Infect Dis 2012;54 Suppl 2:S190-9.

6. Webb C, Ngama M, Ngatia A, et al. Treatment failure among Kenyan children with severe pneumonia--a cohort study. Pediatr Infect Dis J 2012;31:e152-7.

7. Berkley JA, Munywoki P, Ngama M, et al. Viral etiology of severe pneumonia among Kenyan infants and children. JAMA 2010;303:2051-7.

8. Munywoki P. Transmission of Respiratory Syncitial Virus in households: who acquires infection from whom? : Open University; 2013.

9. Munywoki PK, Koech DC, Agoti CN, et al. The source of respiratory syncytial virus infection in infants: a household cohort study in rural Kenya. J Infect Dis 2014;209:1685-92.

10. Hammitt LL, Ojal J, Bashraheil M, et al. Immunogenicity, impact on carriage and reactogenicity of 10-valent pneumococcal non-typeable *Haemophilus influenzae* protein

D conjugate vaccine in Kenyan children aged 1-4 years: a randomized controlled trial. PLoS One 2014;9:e85459.

11. Challenges in global immunization and the Global Immunization Vision and Strategy 2006–2015. Wkly Epidemiol Rec 2006;81:189.

12. Cutts FT, Zaman SM, Enwere G, 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 2005;365:1139-46.

13. Scott JA, Mlacha Z, Nyiro J, et al. Diagnosis of invasive pneumococcal disease among children in Kenya with enzyme-linked immunosorbent assay for immunoglobulin G antibodies to pneumococcal surface adhesin A. Clin Diagn Lab Immunol 2005;12:1195-201.

14. Vernet G, Saha S, Satzke C, et al. Laboratory-based diagnosis of pneumococcal pneumonia: state of the art and unmet needs. Clin Microbiol Infect 2011;17 Suppl 3:1-13.

15. Azzari C, Cortimiglia M, Moriondo M, et al. Pneumococcal DNA is not detectable in the blood of healthy carrier children. J Med Microbiol 2011.

16. Bentley SD, Aanensen DM, Mavroidi A, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS genetics 2006;2:e31.

17. Calix JJ, Nahm MH. A new pneumococcal serotype, 11E, has a variably inactivated wcjE gene. J Infect Dis 2010;202:29-38.

18. Pilishvili T, Lexau C, Farley MM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 2010;201:32-41.

19. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet 2011.

20. 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 2011;11:760-8.

21. Johnson DR, D'Onise K, Holland RA, Raupach JC, Koehler AP. Pneumococcal disease in South Australia: vaccine success but no time for complacency. Vaccine 2012;30:2206-11.

22. Mate C, Omolo L, Karani A, et al. NetSPEAR: Six years of surveillance for pneumococcal disease in the East African Region. 7th International Symposium for Pneumococcus and Pneumococcal Diseases; 2010; Tel Aviv, Israel.

23. Austrian R. Some aspects of the pneumococcal carrier state. J Antimicrob Chemother 1986;18 Suppl A:35-45.

24. Austrian R, Gold J. Pneumococcal Bacteremia with Especial Reference to Bacteremic Pneumococcal Pneumonia. Ann Intern Med 1964;60:759-76.

25. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. J Infect Dis 2003;187:1424-32.

26. Brueggemann AB, Muroki BM, Kulohoma BW, et al. Population genetic structure of *Streptococcus pneumoniae* in Kilifi, Kenya, prior to the introduction of pneumococcal conjugate vaccine. PLoS One 2013;8:e81539.

27. Almeida ST, de Lencastre H, Sa-Leao R. Epidemiology and population structure of serotypes 1, 5 and 7f carried by children in Portugal from 1996-2010 before introduction of the 10-valent and 13-valent pneumococcal conjugate vaccines. PLoS One 2013;8:e75442.

28. Tettelin H, Nelson KE, Paulsen IT, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 2001;293:498-506.

29. Obert C, Sublett J, Kaushal D, et al. Identification of a Candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. Infect Immun 2006;74:4766-77.

30. Embry A, Hinojosa E, Orihuela CJ. Regions of Diversity 8, 9 and 13 contribute to *Streptococcus pneumoniae* virulence. BMC Microbiol 2007;7:80.

31. Wei H, Havarstein LS. Fratricide is essential for efficient gene transfer between pneumococci in biofilms. Applied and environmental microbiology 2012;78:5897-905.

32. Shak JR, Vidal JE, Klugman KP. Influence of bacterial interactions on pneumococcal colonization of the nasopharynx. Trends in microbiology 2013;21:129-35.

33. Kim JO, Weiser JN. Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. J Infect Dis 1998;177:368-77.

34. Sanchez CJ, Hinojosa CA, Shivshankar P, et al. Changes in capsular serotype alter the surface exposure of pneumococcal adhesins and impact virulence. PLoS One 2011;6:e26587.

35. van Rossum AM, Lysenko ES, Weiser JN. Host and bacterial factors contributing to the clearance of colonization by *Streptococcus pneumoniae* in a murine model. Infect Immun 2005;73:7718-26.

36. Kimaro Mlacha SZ, Romero-Steiner S, Hotopp JC, et al. Phenotypic, genomic, and transcriptional characterization of *Streptococcus pneumoniae* interacting with human pharyngeal cells. BMC Genomics 2013;14:383.

37. Price KE, Greene NG, Camilli A. Export requirements of pneumolysin in *Streptococcus pneumoniae*. J Bacteriol 2012;194:3651-60.

38. Lipsitch M, Whitney CG, Zell E, Kaijalainen T, Dagan R, Malley R. Are anticapsular antibodies the primary mechanism of protection against invasive pneumococcal disease? PLoS Med 2005;2:e15.

39. Vernatter J, Pirofski LA. Current concepts in host-microbe interaction leading to pneumococcal pneumonia. Current opinion in infectious diseases 2013;26:277-83.

40. Picard C, von Bernuth H, Ghandil P, et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. Medicine (Baltimore) 2010;89:403-25.

41. Bogaert D, Thompson CM, Trzcinski K, Malley R, Lipsitch M. The role of complement in innate and adaptive immunity to pneumococcal colonization and sepsis in a murine model. Vaccine 2010;28:681-5.

42. Picard C, Puel A, Bustamante J, Ku CL, Casanova JL. Primary immunodeficiencies associated with pneumococcal disease. Current opinion in allergy and clinical immunology 2003;3:451-9.

43. von Gottberg A, Cohen C, de Gouveia L, et al. Epidemiology of invasive pneumococcal disease in the pre-conjugate vaccine era: South Africa, 2003-2008. Vaccine 2013;31:4200-8.

44. Neill DR, Coward WR, Gritzfeld JF, et al. Density and duration of pneumococcal carriage is maintained by transforming growth factor beta1 and T regulatory cells. American journal of respiratory and critical care medicine 2014;189:1250-9.

45. Verschoor CP, Dorrington MG, Novakowski KE, et al. MicroRNA-155 is required for the clearance of *Streptococcus pneumoniae* from the nasopharynx. Infect Immun 2014.

46. Wright AK, Bangert M, Gritzfeld JF, et al. Experimental human pneumococcal carriage augments IL-17A-dependent T-cell defence of the lung. PLoS Pathog 2013;9:e1003274.

47. Das R, LaRose MI, Hergott CB, Leng L, Bucala R, Weiser JN. Macrophage migration inhibitory factor promotes clearance of pneumococcal colonization. Journal of immunology 2014;193:764-72.

48. Joyce EA, Popper SJ, Falkow S. *Streptococcus pneumoniae* nasopharyngeal colonization induces type I interferons and interferon-induced gene expression. BMC Genomics 2009;10:404.

49. Bogaert D, Weinberger D, Thompson C, Lipsitch M, Malley R. Impaired innate and adaptive immunity to *Streptococcus pneumoniae* and its effect on colonization in an infant mouse model. Infect Immun 2009;77:1613-22.

50. Krone CL, Trzcinski K, Zborowski T, Sanders EA, Bogaert D. Impaired innate mucosal immunity in aged mice permits prolonged *Streptococcus pneumoniae* colonization. Infect Immun 2013;81:4615-25.

51. Ferreira DM, Neill DR, Bangert M, et al. Controlled human infection and rechallenge with *Streptococcus pneumoniae* reveals the protective efficacy of carriage in healthy adults. American journal of respiratory and critical care medicine 2013;187:855-64.

52. Cohen JM, Khandavilli S, Camberlein E, Hyams C, Baxendale HE, Brown JS. Protective contributions against invasive *Streptococcus pneumoniae* pneumonia of

antibody and Th17-cell responses to nasopharyngeal colonisation. PLoS One 2011;6:e25558.

53. Malley R, Trzcinski K, Srivastava A, Thompson CM, Anderson PW, Lipsitch M. CD4+ T cells mediate antibody-independent acquired immunity to pneumococcal colonization. Proc Natl Acad Sci U S A 2005;102:4848-53.

54. Lu L, Ma Z, Jokiranta TS, Whitney AR, DeLeo FR, Zhang JR. Species-specific interaction of *Streptococcus pneumoniae* with human complement factor H. Journal of immunology 2008;181:7138-46.

55. Pletz MW, Welte T, Klugman KP. The paradox in pneumococcal serotypes: highly invasive does not mean highly lethal. The European respiratory journal 2010;36:712-3.

56. Stroeher UH, Kidd SP, Stafford SL, Jennings MP, Paton JC, McEwan AG. A pneumococcal MerR-like regulator and S-nitrosoglutathione reductase are required for systemic virulence. J Infect Dis 2007;196:1820-6.

57. Yamaguchi M, Terao Y, Mori-Yamaguchi Y, et al. *Streptococcus pneumoniae* invades erythrocytes and utilizes them to evade human innate immunity. PLoS One 2013;8:e77282.

58. Agarwal V, Asmat TM, Dierdorf NI, Hauck CR, Hammerschmidt S. Polymeric immunoglobulin receptor-mediated invasion of *Streptococcus pneumoniae* into host cells requires a coordinate signaling of SRC family of protein-tyrosine kinases, ERK, and c-Jun N-terminal kinase. The Journal of biological chemistry 2010;285:35615-23.

59. Iovino F, Brouwer MC, van de Beek D, Molema G, Bijlsma JJ. Signalling or binding: the role of the platelet-activating factor receptor in invasive pneumococcal disease. Cellular microbiology 2013;15:870-81.

60. Gradstedt H, Iovino F, Bijlsma JJ. *Streptococcus pneumoniae* invades endothelial host cells via multiple pathways and is killed in a lysosome dependent manner. PLoS One 2013;8:e65626.

61. Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom signaling induces *Streptococcus pneumoniae* biofilm dispersion and transition from asymptomatic colonization to disease. mBio 2013;4.

62. Pettigrew MM, Marks LR, Kong Y, Gent JF, Roche-Hakansson H, Hakansson AP. *Streptococcus pneumoniae* and influenza: dynamic changes in the pneumococcal transcriptome during transition from biofilm formation to invasive disease. Infect Immun 2014.

63. Harvey RM, Hughes CE, Paton AW, Trappetti C, Tweten RK, Paton JC. The impact of pneumolysin on the macrophage response to *Streptococcus pneumoniae* is strain-dependent. PLoS One 2014;9:e103625.

64. Simell B, Auranen K, Kayhty H, Goldblatt D, Dagan R, O'Brien KL. The fundamental link between pneumococcal carriage and disease. Expert review of vaccines 2012;11:841-55.

65. van der Poll T, Opal SM. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet 2009;374:1543-56.

66. Gray BM, Converse GM, 3rd, Dillon HC, Jr. Epidemiologic studies of *Streptococcus pneumoniae* in infants: acquisition, carriage, and infection during the first 24 months of life. J Infect Dis 1980;142:923-33.

67. Bogaert D, De Groot R, Hermans PW. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis 2004;4:144-54.

68. van den Bergh MR, Biesbroek G, Rossen JW, et al. Associations between pathogens in the upper respiratory tract of young children: interplay between viruses and bacteria. PLoS One 2012;7:e47711.

69. Fairchok MP, Martin ET, Chambers S, et al. Epidemiology of viral respiratory tract infections in a prospective cohort of infants and toddlers attending daycare. J Clin Virol 2010;49:16-20.

70. Pappas DE, Hendley JO, Hayden FG, Winther B. Symptom profile of common colds in school-aged children. Pediatr Infect Dis J 2008;27:8-11.

71. Peltola V, Heikkinen T, Ruuskanen O, et al. Temporal association between rhinovirus circulation in the community and invasive pneumococcal disease in children. Pediatr Infect Dis J 2011;30:456-61.

72. Jansen AG, Sanders EA, A VDE, AM VANL, Hoes AW, Hak E. Invasive pneumococcal and meningococcal disease: association with influenza virus and respiratory syncytial virus activity? Epidemiol Infect 2008;136:1448-54.

73. Murdoch DR, Jennings LC. Association of respiratory virus activity and environmental factors with the incidence of invasive pneumococcal disease. J Infect 2009;58:37-46.

74. Wolter N, Cohen C, Tempia S, et al. HIV and influenza virus infections are associated with increased blood pneumococcal load: a prospective, hospital-based observational study in South Africa, 2009-2011. J Infect Dis 2014;209:56-65.

75. O'Brien KL, Walters MI, Sellman J, et al. Severe pneumococcal pneumonia in previously healthy children: the role of preceding influenza infection. Clin Infect Dis 2000;30:784-9.

76. Chertow DS, Memoli MJ. Bacterial coinfection in influenza: a grand rounds review. JAMA 2013;309:275-82.

77. Avadhanula V, Rodriguez CA, Devincenzo JP, et al. Respiratory viruses augment the adhesion of bacterial pathogens to respiratory epithelium in a viral species- and cell type-dependent manner. J Virol 2006;80:1629-36.

78. Golda A, Malek N, Dudek B, et al. Infection with human coronavirus NL63 enhances streptococcal adherence to epithelial cells. J Gen Virol;92:1358-68.

79. Kimaro Mlacha SZ, Peret TC, Kumar N, et al. Transcriptional adaptation of pneumococci and human pharyngeal cells in the presence of a virus infection. BMC Genomics 2013;14:378.

80. Yokota S, Okabayashi T, Hirakawa S, Tsutsumi H, Himi T, Fujii N. Clarithromycin suppresses human respiratory syncytial virus infection-induced *Streptococcus pneumoniae* adhesion and cytokine production in a pulmonary epithelial cell line. Mediators of inflammation 2012;2012:528568.

81. Avadhanula V, Wang Y, Portner A, Adderson E. Nontypeable *Haemophilus influenzae* and *Streptococcus pneumoniae* bind respiratory syncytial virus glycoprotein. J Med Microbiol 2007;56:1133-7.

82. Hament JM, Aerts PC, Fleer A, et al. Direct binding of respiratory syncytial virus to pneumococci: a phenomenon that enhances both pneumococcal adherence to human epithelial cells and pneumococcal invasiveness in a murine model. Pediatric research 2005;58:1198-203.

83. Hament JM, Aerts PC, Fleer A, et al. Enhanced adherence of *Streptococcus pneumoniae* to human epithelial cells infected with respiratory syncytial virus. Pediatric research 2004;55:972-8.

84. Ishizuka S, Yamaya M, Suzuki T, et al. Effects of rhinovirus infection on the adherence of *Streptococcus pneumoniae* to cultured human airway epithelial cells. J Infect Dis 2003;188:1928-39.

85. Wang JH, Kwon HJ, Jang YJ. Rhinovirus enhances various bacterial adhesions to nasal epithelial cells simultaneously. The Laryngoscope 2009;119:1406-11.

86. Trappetti C, Kadioglu A, Carter M, et al. Sialic acid: a preventable signal for pneumococcal biofilm formation, colonization, and invasion of the host. J Infect Dis 2009;199:1497-505.

87. Siegel SJ, Roche AM, Weiser JN. Influenza promotes pneumococcal growth during coinfection by providing host sialylated substrates as a nutrient source. Cell host & microbe 2014;16:55-67.

88. Raza MW, Blackwell CC, Elton RA, Weir DM. Bactericidal activity of a monocytic cell line (THP-1) against common respiratory tract bacterial pathogens is depressed after infection with respiratory syncytial virus. J Med Microbiol 2000;49:227-33.

89. Diavatopoulos DA, Short KR, Price JT, et al. Influenza A virus facilitates *Streptococcus pneumoniae* transmission and disease. FASEB J 2010;24:1789-98.

90. Smith AM, Adler FR, Ribeiro RM, et al. Kinetics of coinfection with influenza A virus and *Streptococcus pneumoniae*. PLoS Pathog 2013;9:e1003238.

91. Mina MJ, Klugman KP, McCullers JA. Live Attenuated Influenza Vaccine, But Not Pneumococcal Conjugate Vaccine, Protects Against Increased Density and Duration of Pneumococcal Carriage After Influenza Infection in Pneumococcal Colonized Mice. J Infect Dis 2013.

92. McCullers JA, McAuley JL, Browall S, Iverson AR, Boyd KL, Henriques Normark B. Influenza enhances susceptibility to natural acquisition of and disease due to *Streptococcus pneumoniae* in ferrets. J Infect Dis 2010;202:1287-95.

93. Short KR, Reading PC, Wang N, Diavatopoulos DA, Wijburg OL. Increased nasopharyngeal bacterial titers and local inflammation facilitate transmission of *Streptococcus pneumoniae.* mBio 2012;3.

94. Wren JT, Blevins LA, Pang B, et al. Influenza A Virus Alters Pneumococcal Nasal Colonization and Middle Ear Infection Independently of Phase Variation. Infect Immun 2014.

95. Stark JM, Stark MA, Colasurdo GN, LeVine AM. Decreased bacterial clearance from the lungs of mice following primary respiratory syncytial virus infection. Journal of medical virology 2006;78:829-38.

96. Kukavica-Ibrulj I, Hamelin ME, Prince GA, et al. Infection with human metapneumovirus predisposes mice to severe pneumococcal pneumonia. J Virol 2009;83:1341-9.

97. Verkaik NJ, Nguyen DT, de Vogel CP, et al. *Streptococcus pneumoniae* exposure is associated with human metapneumovirus seroconversion and increased susceptibility to in vitro HMPV infection. Clin Microbiol Infect 2011;17:1840-4.

98. Rodrigues F, Foster D, Nicoli E, et al. Relationships between rhinitis symptoms, respiratory viral infections and nasopharyngeal colonization with *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* in children attending daycare. Pediatr Infect Dis J 2013;32:227-32.

99. Vu HT, Yoshida LM, Suzuki M, et al. Association between nasopharyngeal load of *Streptococcus pneumoniae*, viral coinfection, and radiologically confirmed pneumonia in Vietnamese children. Pediatr Infect Dis J 2011;30:11-8.

100. Launes C, de-Sevilla MF, Selva L, Garcia-Garcia JJ, Pallares R, Munoz-Almagro C. Viral coinfection in children less than five years old with invasive pneumococcal disease. Pediatr Infect Dis J 2012;31:650-3.

101. Korppi M. Are all pneumococcal infections mixed infections with viruses? Pediatr Infect Dis J 2012;31:1000-1; author reply 1.

102. Adegbola RA, DeAntonio R, Hill PC, et al. Carriage of *Streptococcus pneumoniae* and other respiratory bacterial pathogens in low and lower-middle income countries: a systematic review and meta-analysis. PLoS One 2014;9:e103293.

103. Roca A, Bottomley C, Hill PC, et al. Effect of age and vaccination with a pneumococcal conjugate vaccine on the density of pneumococcal nasopharyngeal carriage. Clin Infect Dis 2012;55:816-24.

104. Gray BM, Turner ME, Dillon HC, Jr. 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 1982;116:692-703.

105. Lipsitch M, Abdullahi O, D'Amour A, et al. Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model. Epidemiology 2012;23:510-9.

106. Dagan R, Shriker O, Hazan I, et al. Prospective study to determine clinical relevance of detection of pneumococcal DNA in sera of children by PCR. J Clin Microbiol 1998;36:669-73.

107. Verhelst R, Kaijalainen T, De Baere T, et al. Comparison of five genotypic techniques for identification of optochin-resistant pneumococcus-like isolates. J Clin Microbiol 2003;41:3521-5.

108. Whatmore AM, Efstratiou A, Pickerill AP, et al. Genetic relationships between clinical isolates of *Streptococcus pneumoniae, Streptococcus oralis*, and *Streptococcus mitis*: characterization of "Atypical" pneumococci and organisms allied to *S. mitis* harboring *S. pneumoniae* virulence factor-encoding genes. Infect Immun 2000;68:1374-82.

109. Harris KA, Turner P, Green EA, Hartley JC. Duplex real-time PCR assay for detection of *Streptococcus pneumoniae* in clinical samples and determination of penicillin susceptibility. J Clin Microbiol 2008;46:2751-8.

110. Seki M, Yamashita Y, Torigoe H, Tsuda H, Sato S, Maeno M. Loop-mediated isothermal amplification method targeting the *lytA* gene for detection of *Streptococcus pneumoniae*. J Clin Microbiol 2005;43:1581-6.

111. Hassan-King M, Baldeh I, Secka O, Falade A, Greenwood B. Detection of *Streptococcus pneumoniae* DNA in blood cultures by PCR. J Clin Microbiol 1994;32:1721-4.

112. Kumar S, Wang L, Fan J, et al. Detection of 11 common viral and bacterial pathogens causing community-acquired pneumonia or sepsis in asymptomatic patients by using a multiplex reverse transcription-PCR assay with manual (enzyme hybridization) or automated (electronic microarray) detection. J Clin Microbiol 2008;46:3063-72.

113. Zhang Y, Isaacman DJ, Wadowsky RM, Rydquist-White J, Post JC, Ehrlich GD. Detection of *Streptococcus pneumoniae* in whole blood by PCR. J Clin Microbiol 1995;33:596-601.

114. Stralin K, Tornqvist E, Kaltoft MS, Olcen P, Holmberg H. Etiologic diagnosis of adult bacterial pneumonia by culture and PCR applied to respiratory tract samples. J Clin Microbiol 2006;44:643-5.

115. Morozumi M, Nakayama E, Iwata S, et al. Simultaneous detection of pathogens in clinical samples from patients with community-acquired pneumonia by real-time PCR with pathogen-specific molecular beacon probes. J Clin Microbiol 2006;44:1440-6.

116. Sheppard CL, Harrison TG, Morris R, Hogan A, George RC. Autolysin-targeted LightCycler assay including internal process control for detection of *Streptococcus pneumoniae* DNA in clinical samples. J Med Microbiol 2004;53:189-95.

117. Isaacman DJ, Zhang Y, Reynolds EA, Ehrlich GD. Accuracy of a polymerase chain reaction-based assay for detection of pneumococcal bacteremia in children. Pediatrics 1998;101:813-6.

118. Stralin K, Backman A, Holmberg H, Fredlund H, Olcen P. Design of a multiplex PCR for *Streptococcus pneumoniae, Haemophilus influenzae, Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* to be used on sputum samples. APMIS 2005;113:99-111.

119. Butler JC, Bosshardt SC, Phelan M, et al. Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. J Infect Dis 2003;187:1416-23.

120. Messmer TO, Whitney CG, Fields BS. Use of polymerase chain reaction to identify pneumococcal infection associated with hemorrhage and shock in two previously healthy young children. Clin Chem 1997;43:930-5.

121. Resti M, Micheli A, Moriondo M, et al. Comparison of the effect of antibiotic treatment on the possibility of diagnosing invasive pneumococcal disease by culture or molecular methods: a prospective, observational study of children and adolescents with proven pneumococcal infection. Clin Ther 2009;31:1266-73.

122. Abdeldaim G, Herrmann B, Molling P, et al. Usefulness of real-time PCR for *lytA*, *ply*, and Spn9802 on plasma samples for the diagnosis of pneumococcal pneumonia. Clin Microbiol Infect 2009.

123. Abdeldaim G, Herrmann B, Korsgaard J, Olcen P, Blomberg J, Stralin K. Is quantitative PCR for the pneumolysin (*ply*) gene useful for detection of pneumococcal lower respiratory tract infection? Clin Microbiol Infect 2009;15:565-70.

124. Avni T, Mansur N, Leibovici L, Paul M. PCR using blood for diagnosis of invasive pneumococcal disease: systematic review and meta-analysis. J Clin Microbiol 2010;48:489-96.

125. Carvalho Mda G, Tondella ML, McCaustland K, et al. Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA. J Clin Microbiol 2007;45:2460-6.

126. Albrich WC, Madhi SA, Adrian PV, et al. Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. Clin Infect Dis 2012;54:601-9.

127. Rudolph KM, Parkinson AJ, Black CM, Mayer LW. Evaluation of polymerase chain reaction for diagnosis of pneumococcal pneumonia. J Clin Microbiol 1993;31:2661-6.

128. Llull D, Lopez R, Garcia E. Characteristic signatures of the *lytA* gene provide a basis for rapid and reliable diagnosis of *Streptococcus pneumoniae* infections. J Clin Microbiol 2006;44:1250-6.

129. Marks LR, Reddinger RM, Hakansson AP. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. mBio 2012;3.

130. Muzzi A, Donati C. Population genetics and evolution of the pan-genome of *Streptococcus pneumoniae*. International journal of medical microbiology : IJMM 2011;301:619-22.

131. Greiner O, Day PJ, Bosshard PP, Imeri F, Altwegg M, Nadal D. Quantitative detection of *Streptococcus pneumoniae* in nasopharyngeal secretions by real-time PCR. J Clin Microbiol 2001;39:3129-34.

132. van Haeften R, Palladino S, Kay I, Keil T, Heath C, Waterer GW. A quantitative LightCycler PCR to detect *Streptococcus pneumoniae* in blood and CSF. Diagn Microbiol Infect Dis 2003;47:407-14.

133. Abdeldaim GM, Stralin K, Olcen P, Blomberg J, Herrmann B. Toward a quantitative DNA-based definition of pneumococcal pneumonia: a comparison of *Streptococcus pneumoniae* target genes, with special reference to the Spn9802 fragment. Diagn Microbiol Infect Dis 2008;60:143-50.

134. McAvin JC, Reilly PA, Roudabush RM, et al. Sensitive and specific method for rapid identification of *Streptococcus pneumoniae* using real-time fluorescence PCR. J Clin Microbiol 2001;39:3446-51.

135. Kee C, Palladino S, Kay I, et al. Feasibility of real-time polymerase chain reaction in whole blood to identify *Streptococcus pneumoniae* in patients with community-acquired pneumonia. Diagn Microbiol Infect Dis 2008;61:72-5.

136. Park HK, Lee SJ, Yoon JW, et al. Identification of the *cpsA* gene as a specific marker for the discrimination of *Streptococcus pneumoniae* from viridans group streptococci. J Med Microbiol 2010;59:1146-52.

137. Zbinden A, Kohler N, Bloemberg GV. *recA*-based PCR assay for accurate differentiation of *Streptococcus pneumoniae* from other viridans streptococci. J Clin Microbiol 2011;49:523-7.

138. Greve T, Moller JK. Accuracy of using the *lytA* gene to distinguish *Streptococcus pneumoniae* from related species. J Med Microbiol 2012;61:478-82.

139. Messmer TO, Sampson JS, Stinson A, Wong B, Carlone GM, Facklam RR. Comparison of four polymerase chain reaction assays for specificity in the identification of *Streptococcus pneumoniae*. Diagn Microbiol Infect Dis 2004;49:249-54.

140. Nagai K, Shibasaki Y, Hasegawa K, et al. Evaluation of PCR primers to screen for *Streptococcus pneumoniae* isolates and beta-lactam resistance, and to detect common macrolide resistance determinants. J Antimicrob Chemother 2001;48:915-8.

141. Kee C, Fatovich DM, Palladino S, et al. Specificity of a quantitative real-time polymerase chain reaction assay for the detection of invasive pneumococcal disease: identifying *Streptococcus pneumoniae* using quantitative polymerase chain reaction. Chest 2010;137:243-4.

142. Azzari C, Moriondo M, Indolfi G, et al. Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. PLoS One 2010;5:e9282.

143. Resti M, Moriondo M, Cortimiglia M, et al. Community-Acquired Bacteremic Pneumococcal Pneumonia in Children: Diagnosis and Serotyping by Real-Time Polymerase Chain Reaction Using Blood Samples. Clin Infect Dis 2010;51:1042-9.

144. Azzari C, Moriondo M, Indolfi G, et al. Molecular detection methods and serotyping performed directly on clinical samples improve diagnostic sensitivity and reveal increased incidence of invasive disease by *Streptococcus pneumoniae* in Italian children. J Med Microbiol 2008;57:1205-12.

145. Rello J, Lisboa T, Lujan M, et al. Severity of pneumococcal pneumonia associated with genomic bacterial load. Chest 2009;136:832-40.

146. Cvitkovic Spik V, Beovic B, Pokorn M, et al. Improvement of pneumococcal pneumonia diagnostics by the use of rt-PCR on plasma and respiratory samples. Scandinavian journal of infectious diseases 2013;45:731-7.

147. Werno AM, Anderson TP, Murdoch DR. Association between pneumococcal load and disease severity in adults with pneumonia. J Med Microbiol 2012;61:1129-35.

148. Rouphael N, Steyn S, Bangert M, et al. Use of 2 pneumococcal common protein real-time polymerase chain reaction assays in healthy children colonized with *Streptococcus pneumoniae*. Diagn Microbiol Infect Dis 2011.

149. Carrol ED, Guiver M, Nkhoma S, et al. High pneumococcal DNA loads are associated with mortality in Malawian children with invasive pneumococcal disease. Pediatr Infect Dis J 2007;26:416-22.

150. Esposito S, Zampiero A, Terranova L, et al. Pneumococcal bacterial load colonization as a marker of mixed infection in children with alveolar community-acquired pneumonia and respiratory syncytial virus or rhinovirus infection. Pediatr Infect Dis J 2013;32:1199-204.

151. Albrich WC, Madhi SA, Adrian PV, et al. Pneumococcal colonisation density: a new marker for disease severity in HIV-infected adults with pneumonia. BMJ open 2014;4:e005953.

152. Wolter N, Tempia S, Cohen C, et al. High nasopharyngeal pneumococcal density, increased by viral coinfection, is associated with invasive pneumococcal pneumonia. J Infect Dis 2014;210:1649-57.

153. Hamano-Hasegawa K, Morozumi M, Nakayama E, et al. Comprehensive detection of causative pathogens using real-time PCR to diagnose pediatric community-acquired pneumonia. J Infect Chemother 2008;14:424-32.

154. Lahti E, Peltola V, Waris M, et al. Induced sputum in the diagnosis of childhood community-acquired pneumonia. Thorax 2009;64:252-7.

155. Stralin K, Herrmann B, Abdeldaim G, Olcen P, Holmberg H, Molling P. Comparison of sputum and nasopharyngeal aspirate samples and of the PCR gene targets *lytA* and Spn9802 for quantitative PCR for rapid detection of pneumococcal pneumonia. J Clin Microbiol 2014;52:83-9.

156. 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 1999;180:1171-6.

157. Cheung YB, Zaman SM, Nsekpong ED, 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 2009;28:990-5.

158. O'Brien KL, Millar EV, Zell ER, et al. Effect of pneumococcal conjugate vaccine on nasopharyngeal colonization among immunized and unimmunized children in a community-randomized trial. J Infect Dis 2007;196:1211-20.

159. Hammitt LL, Akech DO, Morpeth SC, et al. Population effect of 10-valent pneumococcal conjugate vaccine on nasopharyngeal carriage of *Streptococcus pneumoniae* and non-typeable *Haemophilus influenzae* in Kilifi, Kenya: findings from cross-sectional carriage studies. The Lancet Global Health 2014;2:e397-405.

160. Millar EV, Watt JP, Bronsdon MA, et al. Indirect effect of 7-valent pneumococcal conjugate vaccine on pneumococcal colonization among unvaccinated household members. Clin Infect Dis 2008;47:989-96.

161. Hammitt LL, Bruden DL, Butler JC, et al. Indirect effect of conjugate vaccine on adult carriage of *Streptococcus pneumoniae*: an explanation of trends in invasive pneumococcal disease. J Infect Dis 2006;193:1487-94.

162. McCool TL, Weiser JN. Limited role of antibody in clearance of *Streptococcus pneumoniae* in a murine model of colonization. Infect Immun 2004;72:5807-13.

163. Pasinato A, Indolfi G, Marchisio P, et al. Pneumococcal serotype distribution in 1315 nasopharyngeal swabs from a highly vaccinated cohort of Italian children as detected by RT-PCR. Vaccine 2014;32:1375-81.

164. Auranen K, Rinta-Kokko H, Goldblatt D, et al. Colonisation endpoints in *Streptococcus pneumoniae* vaccine trials. Vaccine 2013;32:153-8.

165. da Gloria Carvalho M, Pimenta FC, Jackson D, et al. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. J Clin Microbiol 2010;48:1611-8.

166. Turner P, Hinds J, Turner C, et al. Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. J Clin Microbiol 2011;49:1784-9.

167. Wyllie AL, Chu ML, Schellens MH, et al. *Streptococcus pneumoniae* in Saliva of Dutch Primary School Children. PLoS One 2014;9:e102045.

168. O'Brien KL, Nohynek H. Report from a WHO Working Group: standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*. Pediatr Infect Dis J 2003;22:e1-11.

169. Hendley JO, Sande MA, Stewart PM, Gwaltney JM, Jr. Spread of *Streptococcus pneumoniae* in families. I. Carriage rates and distribution of types. J Infect Dis 1975;132:55-61.

170. Huebner RE, Dagan R, Porath N, Wasas AD, Klugman KP. Lack of utility of serotyping multiple colonies for detection of simultaneous nasopharyngeal carriage of different pneumococcal serotypes. Pediatr Infect Dis J 2000;19:1017-20.

171. Satzke C, Dunne EM, Porter BD, et al. PneuCarriage project: identifying the optimum pneumococcal serotyping method(s), including detection of multiple serotype carriage. International Symposium on Pneumococcus and Pneumococcal Diseases; 2014; Hyderabad, India.

172. Brugger SD, Hathaway LJ, Muhlemann K. Detection of *Streptococcus pneumoniae* strain cocolonization in the nasopharynx. J Clin Microbiol 2009;47:1750-6.

173. Ota MO, Roca A, Bottomley C, et al. Pneumococcal antibody concentrations of subjects in communities fully or partially vaccinated with a seven-valent pneumococcal conjugate vaccine. PLoS One 2012;7:e42997.

174. Gritzfeld JF, Gordon SB, Cremers A. Detection limits in pneumococcal carriage. Pediatr Infect Dis J 2013;32:425-6.

175. Gritzfeld JF, Cremers AJ, Ferwerda G, et al. Density and duration of experimental human pneumococcal carriage. Clin Microbiol Infect 2014.

176. Chien YW, Vidal JE, Grijalva CG, et al. Density interactions among *Streptococcus pneumoniae, Haemophilus influenzae* and *Staphylococcus aureus* in the nasopharynx of young Peruvian children. Pediatr Infect Dis J 2013;32:72-7.

177. Carvalho Mda G, Bigogo GM, Junghae M, et al. Potential nonpneumococcal confounding of PCR-based determination of serotype in carriage. J Clin Microbiol 2012;50:3146-7.

178. Carvalho Mda G, Pimenta FC, Moura I, et al. Non-pneumococcal mitis-group streptococci confound detection of pneumococcal capsular serotype-specific loci in upper respiratory tract. PeerJ 2013;1:e97.

179. Wanjiru E, Foster D, Kosi V, Karani A, Morpeth SC, Scott JA. Molecular serotyping in a quality assurance scheme for the Quellung reaction, pending introduction of the pneumococcal conjugate vaccine to Kenya. 7th International Symposium on Pneumococci and Pneumococcal Diseases; 2010 March 14-18, 2010; Tel Aviv, Israel.

180. Hirama T, Yamaguchi T, Miyazawa H, et al. Prediction of the pathogens that are the cause of pneumonia by the battlefield hypothesis. PLoS One 2011;6:e24474.

181. Alanio A, Desoubeaux G, Sarfati C, et al. Real-time PCR assay-based strategy for differentiation between active *Pneumocystis jirovecii* pneumonia and colonization in immunocompromised patients. Clin Microbiol Infect 2011;17:1531-7.

182. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009;55:611-22.

183. Huggett J, Laver T, Tamisak S, et al. Considerations for the development and application of control materials to improve metagenomic microbial community profiling. Accred Qual Assur 2013;18:77-83.

184. Smith MD, Sheppard CL, Hogan A, et al. Diagnosis of *Streptococcus pneumoniae* infections in adults with bacteremia and community-acquired pneumonia: clinical comparison of pneumococcal PCR and urinary antigen detection. J Clin Microbiol 2009;47:1046-9.

185. Vu HT, Yoshida LM, Suzuki M, et al. Association between nasopharyngeal load of *Streptococcus pneumoniae*, viral coinfection, and radiologically confirmed pneumonia in Vietnamese children. Pediatr Infect Dis J 2010;30:11-8.

186. Smieja M, Castriciano S, Carruthers S, et al. Development and evaluation of a flocked nasal midturbinate swab for self-collection in respiratory virus infection diagnostic testing. J Clin Microbiol 2010;48:3340-2.

187. Bandt D, Monecke S. Development and evaluation of a real-time PCR assay for detection of *Pneumocystis jiroveci.* Transpl Infect Dis 2007;9:196-202.

188. Boddicker JD, Rota PA, Kreman T, et al. Real-time reverse transcription-PCR assay for detection of mumps virus RNA in clinical specimens. J Clin Microbiol 2007;45:2902-8.

189. Ohrmalm L, Wong M, Rotzen-Ostlund M, Norbeck O, Broliden K, Tolfvenstam T. Flocked nasal swab versus nasopharyngeal aspirate for detection of respiratory tract viruses in immunocompromised adults: a matched comparative study. BMC Infect Dis 2010;10:340.

190. Cannas A, Kalunga G, Green C, et al. Implications of storing urinary DNA from different populations for molecular analyses. PLoS One 2009;4:e6985.

191. The mfold Web Server. Rensselaer Polytechnic Institute. at http://mfold.rna.albany.edu/?q=mfold/dna-folding-form.

192. BuG@S. at http://bugs.sgul.ac.uk/.

193. Said MA, Johnson HL, Nonyane BA, et al. Estimating the burden of pneumococcal pneumonia among adults: a systematic review and meta-analysis of diagnostic techniques. PLoS One 2013;8:e60273.

194. Charkaluk ML, Kalach N, Mvogo H, et al. Assessment of a rapid urinary antigen detection by an immunochromatographic test for diagnosis of pneumococcal infection in children. Diagn Microbiol Infect Dis 2006;55:89-94.

195. Abdullahi O, Karani A, Tigoi CC, et al. The prevalence and risk factors for pneumococcal colonization of the nasopharynx among children in Kilifi District, Kenya. PLoS One 2012;7:e30787.

196. Scott JA, Bauni E, Moisi JC, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). Int J Epidemiol 2012;41:650-7.

197. Chow SC, Shao J, Wang H. Sample Size Calculations in Clinical Research. 2nd ed. New York: Marcel Dekker; 2008.

198. Louria DB, Blumenfeld HL, Ellis JT, Kilbourne ED, Rogers DE. Studies on influenza in the pandemic of 1957-1958. II. Pulmonary complications of influenza. J Clin Invest 1959;38:213-65.

199. Grijalva CG, Griffin MR, Edwards KM, et al. The role of influenza and parainfluenza infections in nasopharyngeal pneumococcal acquisition among young children. Clin Infect Dis 2014;58:1369-76.

200. Wisdom A, Leitch EC, Gaunt E, Harvala H, Simmonds P. Screening respiratory samples for detection of human rhinoviruses (HRVs) and enteroviruses: comprehensive VP4-VP2 typing reveals high incidence and genetic diversity of HRV species C. J Clin Microbiol 2009;47:3958-67.

201. Hammitt LL, Kazungu S, Welch S, et al. Added value of an oropharyngeal swab in detection of viruses in children hospitalized with lower respiratory tract infection. J Clin Microbiol 2011;49:2318-20.

202. Gwaltney JM, Jr., Sande MA, Austrian R, Hendley JO. Spread of *Streptococcus pneumoniae* in families. II. Relation of transfer of *S. pneumoniae* to incidence of colds and serum antibody. J Infect Dis 1975;132:62-8.

203. von Gottberg A, de Gouveia L, Tempia S, et al. Effects of vaccination on invasive pneumococcal disease in South Africa. N Engl J Med 2014;371:1889-99.

204. Tigoi CC, Gatakaa H, Karani A, et al. Rates of acquisition of pneumococcal colonization and transmission probabilities, by serotype, among newborn infants in Kilifi District, Kenya. Clin Infect Dis 2012;55:180-8.

205. El Aila NA, Emler S, Kaijalainen T, et al. The development of a 16S rRNA gene based PCR for the identification of *Streptococcus pneumoniae* and comparison with four other species specific PCR assays. BMC Infect Dis 2010;10:104.

206. Dunne EM, Manning J, Russell FM, Robins-Browne RM, Mulholland EK, Satzke C. Effect of pneumococcal vaccination on nasopharyngeal carriage of *Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis*, and *Staphylococcus aureus* in Fijian children. J Clin Microbiol 2012;50:1034-8.

207. Adetifa IM, Antonio M, Okoromah CA, et al. Pre-vaccination nasopharyngeal pneumococcal carriage in a Nigerian population: epidemiology and population biology. PLoS One 2012;7:e30548.

208. Hill PC, Cheung YB, Akisanya A, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian infants: a longitudinal study. Clin Infect Dis 2008;46:807-14.

209. Abdullahi O, Wanjiru E, Musyimi R, Glass N, Scott JA. Validation of nasopharyngeal sampling and culture techniques for detection of *Streptococcus pneumoniae* in children in Kenya. J Clin Microbiol 2007;45:3408-10.

210. Weddle G, Jackson MA, Selvarangan R. Reducing blood culture contamination in a pediatric emergency department. Pediatr Emerg Care 2011;27:179-81.

211. Murdoch DR, Anderson TP, Beynon KA, et al. Evaluation of a PCR assay for detection of *Streptococcus pneumoniae* in respiratory and nonrespiratory samples from adults with community-acquired pneumonia. J Clin Microbiol 2003;41:63-6.

Chapter Fourteen

14 Appendices

14.1 Appendix A: STATA code for selection of viral episodes

14.1.1 Cleaning the data and generating extra variables needed

The STATA code began with removal of unnecessary variables and specimens other than nasopharyngeal swabs, and checking for duplicates in terms of study ID and sample date, which together made up the unique identifier for a nasopharyngeal swab in this dataset. The study ID was unique to the participant. Variables such as number of children per household, number of smokers per household, poverty level, cooking fuel type, cooking inside the house, type of toilet available, water source; that could potentially be associated with nasopharyngeal pneumococcal carriage were generated. All swabs from participants over the age of five years at enrolment were then dropped from the data. Five households had rhinovirus subtyping data available, but with different variable names, these were harmonised under one variable name. Any Ct values that were improbably low were checked with the laboratory team and changed to 0, as these were due to wandering baseline, and not true amplification.

The nasopharyngeal swab PCR data was made binary for each viral target, with a Ct value of ≤35 counted as positive, and a Ct>35, an undetected Ct or a missing Ct counted as negative. Each rhinovirus subtype was considered as an individual viral target.

An upper respiratory tract infection (URTI) was defined as the presence of cough, coryza, sore throat, nasal flaring or wheeze, in practice, all URTI were made up of a cough or coryza or both.

clear matrix

clear

set mem 500m

set more off

capture log close

log using "WAIFW selections", text replace

cd "/Users/smorpeth/Documents/Susan's stuff/Kilifi/PhD Kilifi/WAIFW in PhD"

use "Entire WAIFW dataset/waifw_field_mpx_clean_300412_susan_confounders_rhino5_51_34_19_40.dta"

br

drop location slname village student occupation reasonnfs awaynfs comments awaycomment away anyothercomplains othercomplaints unablefeed headnodding consciouslevel unabletalk cyanosis referralby clinician temp resprate oxygensats heartrate

diag2 diag fw extractiondate extractionby pcrdate pcrby nfss sampledate1 age_code lad_id r_extractiondate r_extractionby serial month dash sampleid extractiondate2 lab_id r_pcrdate r_pcrby sampledate2 sampledateyear sampledatemonth sampledateday sampledateday1 sampledateyear1

drop v9 hhname hhheadname

duplicates report visitid

duplicates report studyid sampledate

****check whether can use this compound unique id without dropping missing or duplicate visitids****

codebook sampletype

drop if sampletype == 2

****2 is oral fluid****

drop if nfs == 0

duplicates report visitid

****Generating variables needed as potential confounders later*******

****Number of children in the household****

```
sort hhid studyid sampledate

egen tag_studyid=tag(studyid)

*this tags the first of each studyid as 1 and all the rest as 0

        tab tag_studyid,

        codebook studyid



by hhid:  gen nkids=sum(tag_studyid) if age_start<16

by hhid: replace nkids=nkids[_n-1] if age_start>=16

by hhid: gen Nkids=nkids[_N]

drop nkids

tab Nkids



***Number of adults in the household who are smokers***

bysort hhid: gen nsmoke=sum(tag_studyid) if smoker==1

by hhid: replace nsmoke=nsmoke[_n-1] if smoker==0 | smoker==.

by hhid: gen Nsmoke=nsmoke[_N]

        replace Nsmoke=0 if Nsmoke==.
```

drop nsmoke

tab Nsmoke

****Poverty level by assets****

codebook bicycle tvvideo motorcycle radio mobilephone electricity cowsbin goatsbin chickenbin turkeybin ducksbin housetype mudwall houseowner

gen assets = bicycle + tvvideo + motorcycle + radio + mobilephone + electricity + cowsbin ///

+ goatsbin + chickenbin + turkeybin + ducksbin

tab assets

****Cooking fuel****

codebook toilet toilettype burnwaste fuelcook placecook watersource watersite

****toilet is any toilet, placecook gives more info than fuelcook, watersource more useful than *watersite*

gen hygiene=1 if toilet==1 & (watersource==1 | watersource==4)

**watersource 1 is a closed well, 4 is piped. Other watersources are open well, open and closed well, *piped and open well or missing*

replace hygiene=2 if toilet==1 & watersource!=1 & watersource!=4

replace hygiene=3 if toilet==0 & (watersource==1 | watersource==4)

replace hygiene=4 if toilet==0 & watersource!=1 & watersource!=4

replace hygiene=. if toilet==. | watersource==.

tab hygiene

****1 is the best hygiene score (has a toilet and a safe water source), 4 is the worst (has neither), toilet *scores as more important than the water source*

tab placecook

gen cookinside = 3 if placecook==3

replace cookinside = 2 if placecook==1

replace cookinside = 1 if placecook==2

replace cookinside = . if placecook==.

tab cookinside

****1 is the best cookinside score (cooks outside), 2 is moderate (cooks in different house), 3 is worst (cooks in house where sleeps)*

drop if age_start > 5

duplicates report visitid

```
duplicates list visitid

duplicates report studyid sampledate



replace rhinospp = hrv_species if rhinospp == "" & hrv_species != ""



***Improbably low Ct values***

list studyid sampledate visitid e229 if e229>0 & e229<14

list studyid sampledate visitid nl63 if nl63>0 & nl63<14

 list studyid sampledate visitid oc43 if oc43>0 & oc43<14

 list studyid sampledate visitid adeno if adeno>0 & adeno<14

 list studyid sampledate visitid piv4 if piv4>0 & piv4<14

  list studyid sampledate visitid rhino if rhino>0 & rhino<14

   list studyid sampledate visitid flua flub fluc if (flua>0 & flua<14) | (flub>0 & flub<14) ///

 | (fluc>0 & fluc<14)

*none of the other targets had any Ct<14

replace e229=0 if e229>0 & e229<10

replace nl63=0 if nl63>0 & nl63<10
```

replace oc43=0 if oc43>0 & oc43<10

replace adeno=0 if adeno>0 & adeno<10

replace piv4=0 if piv4>0 & piv4<14

replace rhino=0 if rhino>0 & rhino<14

replace flua=0 if flua>0 & flua<14

replace flub=0 if flub>0 & flub<14

replace fluc=0 if fluc>0 & fluc<14

****RSV-A****

****

gen rsvabin = 1 if rsva > 0 & rsva < 36

replace rsvabin = 0 if rsva == 0

replace rsvabin = 0 if rsva > 35 & rsva < .

replace rsvabin = 0 if rsva == .

****making the Ct values binary ie pos or neg****

****RSV-B****

****

gen rsvbbin = 1 if rsvb > 0 & rsvb < 36

replace rsvbbin = 0 if rsvb == 0

replace rsvbbin = 0 if rsvb > 35 & rsvb < .

replace rsvbbin = 0 if rsvb == .

****making the Ct values binary ie pos or neg****

****OC-43***Coronavirus****

****

gen oc43bin = 1 if oc43 > 0 & oc43 < 36

replace oc43bin = 0 if oc43 == 0

replace oc43bin = 0 if oc43 > 35 & oc43 < .

replace oc43bin = 0 if oc43 == .

****making the Ct values binary ie pos or neg****

****NL-63***Coronavirus****

```
**********

gen nl63bin = 1 if nl63 > 0 & nl63 < 36

replace nl63bin = 0 if nl63 == 0

replace nl63bin = 0 if nl63 > 35 & nl63 < .

replace nl63bin = 0 if nl63 == .

***making the Ct values binary ie pos or neg***



***E-229***Coronavirus***

**********

gen e229bin = 1 if e229 > 0 & e229 < 36

replace e229bin = 0 if e229 == 0

replace e229bin = 0 if e229 > 35 & e229 < .

replace e229bin = 0 if e229 == .

***making the Ct values binary ie pos or neg***



***Rhinovirus A1***

**********
```

gen rhinospp_a1bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "A1"

replace rhinospp_a1bin = 0 if rhino == 0

replace rhinospp_a1bin = 0 if rhino > 35 & rhino < .

replace rhinospp_a1bin = 0 if rhinospp != "A1"

****making the Ct values binary ie pos or neg****

****Rhinovirus A2****

****

gen rhinospp_a2bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "A2"

replace rhinospp_a2bin = 0 if rhino == 0

replace rhinospp_a2bin = 0 if rhino > 35 & rhino < .

replace rhinospp_a2bin = 0 if rhinospp != "A2"

****making the Ct values binary ie pos or neg****

****Rhinovirus A3****

****

gen rhinospp_a3bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "A3"

replace rhinospp_a3bin = 0 if rhino == 0

replace rhinospp_a3bin = 0 if rhino > 35 & rhino < .

replace rhinospp_a3bin = 0 if rhinospp != "A3"

****making the Ct values binary ie pos or neg****

****Rhinovirus A4*********

****

****check in main dataset whether rhinospp or hrv_species contains all of this data for the 4 rhino typed households*

gen rhinospp_a4bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "A4"

replace rhinospp_a4bin = 0 if rhino == 0

replace rhinospp_a4bin = 0 if rhino > 35 & rhino < .

replace rhinospp_a4bin = 0 if rhinospp != "A4"

****making the Ct values binary ie pos or neg****

****Rhinovirus A5*********

****

****check in main dataset whether rhinospp or hrv_species contains all of this data for the 4 rhino typed households*

gen rhinospp_a5bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "A5"

replace rhinospp_a5bin = 0 if rhino == 0

replace rhinospp_a5bin = 0 if rhino > 35 & rhino < .

replace rhinospp_a5bin = 0 if rhinospp != "A5"

****making the Ct values binary ie pos or neg****

****Rhinovirus B1****

****

gen rhinospp_b1bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "B1"

replace rhinospp_b1bin = 0 if rhino == 0

replace rhinospp_b1bin = 0 if rhino > 35 & rhino < .

replace rhinospp_b1bin = 0 if rhinospp != "B1"

****making the Ct values binary ie pos or neg****

****Rhinovirus C1****

```
**********

gen rhinospp_c1bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "C1"

replace rhinospp_c1bin = 0 if rhino == 0

replace rhinospp_c1bin = 0 if rhino > 35 & rhino < .

replace rhinospp_c1bin = 0 if rhinospp != "C1"

***making the Ct values binary ie pos or neg***



***Rhinovirus C2***

**********

gen rhinospp_c2bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "C2"

replace rhinospp_c2bin = 0 if rhino == 0

replace rhinospp_c2bin = 0 if rhino > 35 & rhino < .

replace rhinospp_c2bin = 0 if rhinospp != "C2"

***making the Ct values binary ie pos or neg***



***Rhinovirus C3***

**********
```

gen rhinospp_c3bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "C3"

replace rhinospp_c3bin = 0 if rhino == 0

replace rhinospp_c3bin = 0 if rhino > 35 & rhino < .

replace rhinospp_c3bin = 0 if rhinospp != "C3"

****making the Ct values binary ie pos or neg****

****Rhinovirus C4****

****

gen rhinospp_c4bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "C4"

replace rhinospp_c4bin = 0 if rhino == 0

replace rhinospp_c4bin = 0 if rhino > 35 & rhino < .

replace rhinospp_c4bin = 0 if rhinospp != "C4"

****making the Ct values binary ie pos or neg****

****Rhinovirus C5****

****

gen rhinospp_c5bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "C5"

replace rhinospp_c5bin = 0 if rhino == 0

replace rhinospp_c5bin = 0 if rhino > 35 & rhino < .

replace rhinospp_c5bin = 0 if rhinospp != "C5"

****making the Ct values binary ie pos or neg****

****Rhinovirus D1****

****

gen rhinospp_d1bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "D1"

replace rhinospp_d1bin = 0 if rhino == 0

replace rhinospp_d1bin = 0 if rhino > 35 & rhino < .

replace rhinospp_d1bin = 0 if rhinospp != "D1"

****making the Ct values binary ie pos or neg****

****Rhinovirus E1****

****

gen rhinospp_e1bin = 1 if rhino > 0 & rhino < 36 & rhinospp == "E1"

replace rhinospp_e1bin = 0 if rhino == 0

replace rhinospp_e1bin = 0 if rhino > 35 & rhino < .

replace rhinospp_e1bin = 0 if rhinospp != "E1"

****making the Ct values binary ie pos or neg****

****GENERAL RHINOVIRUS****

gen rhino_bin = 1 if rhino > 0 & rhino < 36

replace rhino_bin = 0 if rhino == 0

replace rhino_bin = 0 if rhino > 35 & rhino < .

replace rhino_bin = 0 if rhino == .

****making the Ct values binary ie pos or neg for general rhino because needed for coinfection assessment****

****UNKNOWN RHINOVIRUS****

gen rhino_Ubin = 1 if rhino > 0 & rhino < 36 & rhinospp =="U"

replace rhino_Ubin = 0 if rhino == 0

replace rhino_Ubin = 0 if rhino > 35 & rhino < .

replace rhino_Ubin = 0 if rhinospp != "U"

****making the Ct values binary ie pos or neg for unknown rhinos because needed for coinfection assessment****

****Influenza A****

****

gen fluabin = 1 if flua > 0 & flua < 36

replace fluabin = 0 if flua == 0

replace fluabin = 0 if flua > 35 & flua < .

replace fluabin = 0 if flua == .

****making the Ct values binary ie pos or neg****

****Influenza B****

****

gen flubbin = 1 if flub > 0 & flub < 36

replace flubbin = 0 if flub == 0

replace flubbin = 0 if flub > 35 & flub < .

replace flubbin = 0 if flub == .

```
***making the Ct values binary ie pos or neg***



***Influenza C***

**********

gen flucbin = 1 if fluc > 0 & fluc < 36

replace flucbin = 0 if fluc == 0

replace flucbin = 0 if fluc > 35 & fluc < .

replace flucbin = 0 if fluc == .

***making the Ct values binary ie pos or neg***


***HMPV***

**********

gen hmpvbin = 1 if hmpv > 0 & hmpv < 36

replace hmpvbin = 0 if hmpv == 0

replace hmpvbin = 0 if hmpv > 35 & hmpv < .

replace hmpvbin = 0 if hmpv == .

***making the Ct values binary ie pos or neg***
```

****Adenovirus****

****

gen adenobin = 1 if adeno > 0 & adeno < 36

replace adenobin = 0 if adeno == 0

replace adenobin = 0 if adeno > 35 & adeno < .

replace adenobin = 0 if adeno == .

****making adeno Ct values binary ie pos or neg, because will need these later to assess co-infections*

****Mycoplasma pneumoniae****

****

gen mycopbin = 1 if mycop > 0 & mycop < 36

replace mycopbin = 0 if mycop == 0

replace mycopbin = 0 if mycop > 35 & mycop < .

replace mycopbin = 0 if mycop == .

****making mycop Ct values binary ie pos or neg, because will need these later to assess co-infections*

****PIV-1****

****

gen piv1bin = 1 if piv1 > 0 & piv1 < 36

replace piv1bin = 0 if piv1 == 0

replace piv1bin = 0 if piv1 > 35 & piv1 < .

replace piv1bin = 0 if piv1 == .

****making piv1 Ct values binary ie pos or neg, because will need these later to assess co-infections*

****PIV-2****

****

gen piv2bin = 1 if piv2 > 0 & piv2 < 36

replace piv2bin = 0 if piv2 == 0

replace piv2bin = 0 if piv2 > 35 & piv2 < .

replace piv2bin = 0 if piv2 == .

****making piv2 Ct values binary ie pos or neg, because will need these later to assess co-infections*

****PIV-3****

****

gen piv3bin = 1 if piv3 > 0 & piv3 < 36

replace piv3bin = 0 if piv3 == 0

replace piv3bin = 0 if piv3 > 35 & piv3 < .

replace piv3bin = 0 if piv3 == .

****making piv3 Ct values binary ie pos or neg, because will need these later to assess co-infections*

****PIV-4****

****

gen piv4bin = 1 if piv4 > 0 & piv4 < 36

replace piv4bin = 0 if piv4 == 0

replace piv4bin = 0 if piv4 > 35 & piv4 < .

replace piv4bin = 0 if piv4 == .

****making piv4 Ct values binary ie pos or neg, because will need these later to assess co-infections*

save "WAIFW_selections_30April12.dta", replace

drop urti

****existing definition of urti is as below but also seems to include "sick" which adds some subjects sick *with other illnesses*

gen urti=cough== 1 | runnynose == 1 | sorethroat == 1

tab urti

tab urti cough

tab urti runnynose

tab cough runnynose

tab urti sorethroat

gen any_virus=0

replace any_virus=1 if rsvabin==1 | rsvbbin==1 | oc43bin==1 | nl63bin==1 | e229bin==1 ///

| fluabin==1 | flubbin==1 | flucbin==1 | hmpvbin==1 | rhino_bin==1

14.1.2 Defining episodes of infection with each viral target

The data was sorted by study ID and sample date, and received an ordered number, within each study ID. The data were then made panel data using the xtset command. A variable for the episode of infection with the first viral target, RSV A, was generated using the binary variable for RSV as defined by the Ct for RSV as above. The episode of infection was then expanded to include all negative swabs that lay between positive swabs as long as those positive swabs were within 14 days of each other, using lag and lead commands in the panel data. Finally if there were not at least 2 positive swabs within a 14-day period, then the episode variable became negative. This ensured that the definition of an episode must include at least 2 swabs positive for RSV-A with a Ct≤35, within a 14-day period.

///define rsvainfection episodes

sort studyid sampledate

by studyid: gen order=_n

xtset studyid order

*rsva episode

****note f means forward and l means lag. see help xtset for a description of these commands***

gen rsvainf=1 if rsvabin==1

replace rsvainf=1 if l.rsvabin==1 & f.rsvabin==1 & rsvainf==. & f.sampledate-l.sampledate<15

replace rsvainf=1 if l.rsvabin==1 & f2.rsvabin==1 & rsvainf==. & f2.sampledate-l.sampledate<15

replace rsvainf=1 if l2.rsvabin==1 & f.rsvabin==1 & rsvainf==. & f.sampledate-l2.sampledate<15

replace rsvainf=1 if l.rsvabin==1 & f3.rsvabin==1 & rsvainf==. & f3.sampledate-l.sampledate<15

replace rsvainf=1 if l2.rsvabin==1 & f2.rsvabin==1 & rsvainf==. & f2.sampledate-l2.sampledate<15

replace rsvainf=1 if l3.rsvabin==1 & f1.rsvabin==1 & rsvainf==. & f1.sampledate-l3.sampledate<15

replace rsvainf=0 if rsvainf==.

replace rsvainf=0 if rsvabin==1 & l.rsvainf==0 & f.rsvainf==0 //must have >1 swab positive

label var rsvainf "rsva episode"

****this codes a swab as part of an rsva episode if the child had at least 2 positive swabs within 2 *weeks*

Next the URTI variable was used to generate a variable for "ever URTI" ie. had the participant had URTI symptoms at any point during that viral infection episode? This was because URTI symptoms can come and go, or be variably reported, so an episode was now defined as; at least 2 swabs positive for RSV-A with a Ct≤35, within a 14-day period, during which the participant suffered from URTI symptoms on at least one occasion. These episodes are marked with a variable indicating the start of an RSV infection episode.

```
by studyid: gen nrsvainf=sum(rsvainf)
by studyid: gen nurti=sum(urti) if rsvainf==1
    by studyid: replace nurti=nurti[_n-1] if nurti==.
    by studyid: gen everurti=nurti[_N]>0

 //start of rsva episodes of interest
gen srsvainf=rsvainf==1 & nrsvainf==1 & everurti==1
 tab srsvainf
```

****this means that only episodes during which the child was symptomatic at any time are marked*

****srsvainf is start of rsva infection episode*

The start of each RSV infection episode was then numbered, because we wanted to be able to select only the first episode in each household. Variables for the number of RSV-A infections per household and the start of the RSV-A infections in that household were generated. If more than one participant per household started an RSV-A infection episode on the same date, then all of those participants were marked as the start of RSV-A infection episodes for that household. All swabs in those first RSV-A episodes per household were marked using the 'nrsvaselect' variable.

by studyid: gen nsrsvainf=sum(srsvainf)

tab nsrsvainf

replace srsvainf=0 if nsrsvainf>1

****nsrsvainf is numbering the start of the rsva infection episodes so that we end up with only the first *episode*

drop nurti everurti

so that it can be used again for the next virus

//identify the first child infected with symptomatic rsva per hh

gsort hhid sampledate -srsvainf studyid

//starts with the rsvainf infections sample were collected the same date

by hhid: gen nrsva_hh=sum(srsvainf)

***nrsva_hh is number of rsva infections in the household

***srsvainf_hh is start of rsva infections in the household

by hhid: gen srsvainf_hh=nrsva_hh==1 & srsvainf==1

by hhid: replace srsvainf_hh=1 if nrsva_hh==2 & nrsva_hh[_n-1]==1 ///

& sampledate==sampledate[_n-1]

by hhid: replace srsvainf_hh=1 if nrsva_hh==3 & nrsva_hh[_n-1]==2 ///

& nrsva_hh==1 & sampledate==sampledate[_n-1]

***ensures that if 2 or 3 children all started with rsva infections on the same date, they are all *included

sort studyid sampledate

by studyid: gen nrsvaselect=sum(srsvainf_hh==1) if rsvainf==1 & nsrsvainf==1

by studyid: gen srsvaselect=1 if nrsvaselect==1 & nrsvaselect[_n-1]!=1

drop nrsvaselect

by studyid: gen nrsvaselect=sum(srsvaselect==1)

replace nrsvaselect=0 if rsvainf==0

****selects the rsva episode samples during the infection episode, if the first for that household*

The data being in order of study ID and sample date, the xtset command was used again to create panel data, again by order within study ID. All swabs within 2 weeks prior to the RSV-A infection episode were marked as "Before" swabs, all swabs during the episode were marked as "During" swabs, and all swabs within 4 weeks after the episode were marked as "After" swabs, using the 'rsvaselect' variable. The number of selected RSV-A episodes and the number of swabs in the "Before", "During" and "After" categories for RSVA were checked using the 'tab' commands.

drop order

by studyid: gen order=_n

xtset studyid order

//select samples

gen rsvaselect=0

//selects samples 2 weeks BEFORE the rsva episode

replace rsvaselect=1 if f.nrsvaselect==1 & f.sampledate-sampledate<15 & nrsvaselect==0

replace rsvaselect=1 if f2.nrsvaselect==1 & f2.sampledate-sampledate<15 & nrsvaselect==0

replace rsvaselect=1 if f3.nrsvaselect==1 & f3.sampledate-sampledate<15 & nrsvaselect==0

replace rsvaselect=1 if f4.nrsvaselect==1 & f4.sampledate-sampledate<15 & nrsvaselect==0

replace rsvaselect=1 if f5.nrsvaselect==1 & f5.sampledate-sampledate<15 & nrsvaselect==0

replace rsvaselect=1 if f6.nrsvaselect==1 & f6.sampledate-sampledate<15 & nrsvaselect==0

//selects samples DURING the rsva episode

replace rsvaselect=2 if nrsvaselect==1

//selects 4 weeks AFTER

replace rsvaselect=3 if l.nrsvaselect==1 & sampledate-l.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l2.nrsvaselect==1 & sampledate-l2.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l3.nrsvaselect==1 & sampledate-l3.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l4.nrsvaselect==1 & sampledate-l4.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l5.nrsvaselect==1 & sampledate-l5.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l6.nrsvaselect==1 & sampledate-l6.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l7.nrsvaselect==1 & sampledate-l7.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l8.nrsvaselect==1 & sampledate-l8.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l9.nrsvaselect==1 & sampledate-l9.sampledate<29 & nrsvaselect==0

replace rsvaselect=3 if l10.nrsvaselect==1 & sampledate-l10.sampledate<29 & nrsvaselect==0

tab srsvainf_hh

tab rsvaselect

A co-infection variable was created so that if a swab selected as being before, during or after an RSV-A episode, was also positive for another virus (or *Mycoplasma pneumoniae*), it was tagged. All of this code was subsequently repeated for each other viral target.

**Make a coinfection variable which is positive if before, during or after any selected viral episode*

**a swab is positive for any other virus*

gen coinfect=0

replace coinfect=1 if (rsvaselect==1 | rsvaselect==2 | rsvaselect==3) & ///

(adenobin==1 | rsvbbin==1 | oc43bin==1 | nl63bin==1 | e229bin==1 | fluabin==1 |

flubbin==1 | flucbin==1 | hmpvbin==1 ///

| rhino_bin==1 | piv1bin==1 | piv2bin==1 | piv3bin==1 | piv4bin==1 | mycopbin==1)

14.1.3 Overall swab selections for all viral targets

Finally, the overall selections for all viruses were marked so that all "Before" swabs were marked, all "During" swabs were marked and all "After" swabs were marked. These selected swabs were exported using the 'outsheet' command, located in the freezer archive, and *lytA* qPCR was run on the DNA previously extracted from the selected swabs.

****Overall Selections****

gen overallselect = 1 if rsvaselect == 1 | rsvbselect == 1 | oc43select == 1 | nl63select == 1

| ///

e229select == 1 | rhinospp_a1select == 1 | rhinospp_a2select == 1 | rhinospp_a3select ==

1 ///

| rhinospp_a4select == 1 | rhinospp_a5select == 1 | rhinospp_b1select == 1 ///

| rhinospp_c1select == 1 | rhinospp_c2select == 1 | rhinospp_c3select == 1 ///

| rhinospp_c4select == 1 | rhinospp_c5select == 1 | rhinospp_d1select == 1 ///

| rhinospp_e1select == 1 | fluaselect == 1 | flubselect == 1 | flucselect == 1 | hmpvselect == 1

replace overallselect = 2 if overallselect == . & (rsvaselect == 2 | rsvbselect == 2 ///

| oc43select == 2 | nl63select == 2 | e229select == 2 | rhinospp_a1select == 2 ///

| rhinospp_a2select == 2 | rhinospp_a3select == 2 | rhinospp_a4select == 2 ///

| rhinospp_a5select == 2 | rhinospp_b1select == 2 | rhinospp_c1select == 2 ///

| rhinospp_c2select == 2 | rhinospp_c3select == 2 | rhinospp_c4select == 2 ///

| rhinospp_c5select == 2 | rhinospp_d1select == 2 | rhinospp_e1select == 2 | fluaselect == 2 ///

| flubselect == 2 | flucselect == 2 | hmpvselect == 2)

replace overallselect = 3 if overallselect == . & (rsvaselect == 3 | rsvbselect == 3 | oc43select == 3 ///

| nl63select == 3 | e229select == 3 | rhinospp_a1select == 3 | rhinospp_a2select == 3 ///

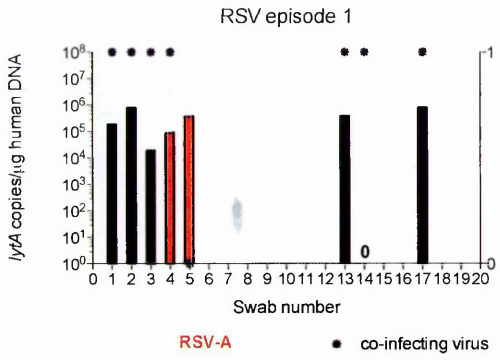
| rhinospp_a3select == 3 | rhinospp_a4select == 3 | rhinospp_a5select == 3 ///

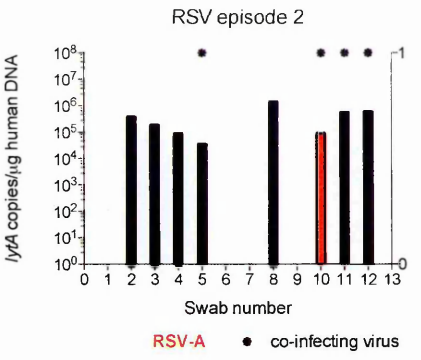
| rhinospp_b1select == 3 | rhinospp_c1select == 3 | rhinospp_c2select == 3 ///

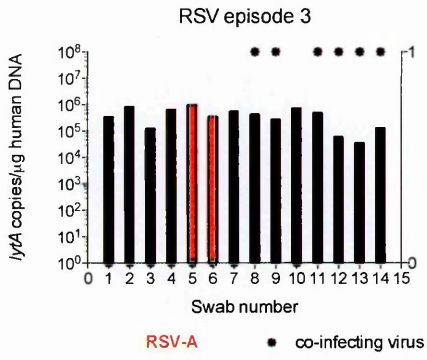
| rhinospp_c3select == 3 | rhinospp_c4select == 3 | rhinospp_c5select == 3 ///

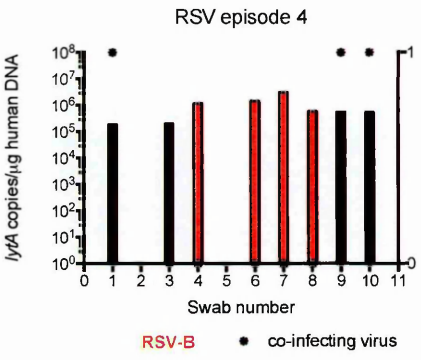
```
| rhinospp_d1select == 3 | rhinospp_e1select == 3 | fluaselect == 3 | flubselect == 3 ///

| flucselect == 3 | hmpvselect == 3)



replace overallselect = 0 if overallselect == .



preserve

keep hhid studyid visitid sampledate overallselect rsvaselect rsvbselect oc43select ///

nl63select e229select rhinospp_a1select rhinospp_a2select rhinospp_a3select ///

rhinospp_a4select rhinospp_a5select rhinospp_b1select rhinospp_c1select ///

rhinospp_c2select rhinospp_c3select rhinospp_c4select rhinospp_c5select ///

rhinospp_d1select rhinospp_e1select fluaselect flubselect flucselect hmpvselect

keep if overallselect != 0

outsheet using "overall_selections_revised300412_susan3_confounders", replace

restore



save "WAIFW_selections_30April12.dta", replace

log close
```

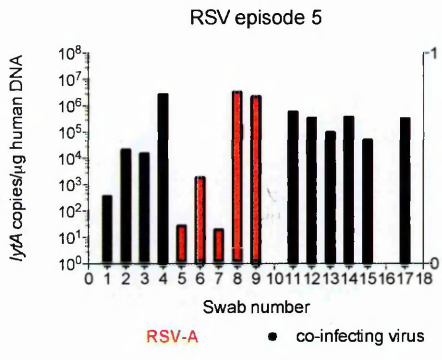
14.2 Appendix B: Nasopharyngeal pneumococcal concentration before, during and after episodes of symptomatic viral upper respiratory tract infection with RSV or rhinovirus

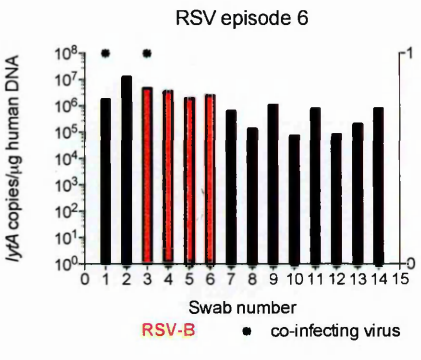

RSV episode 1
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

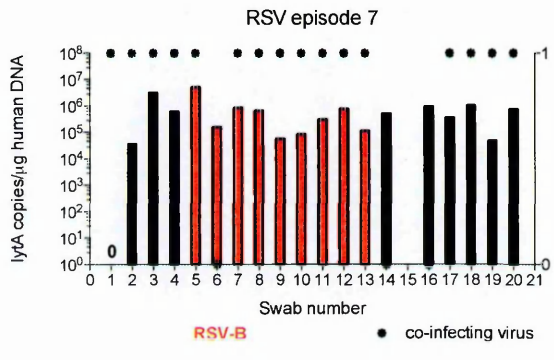

RSV episode 2
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

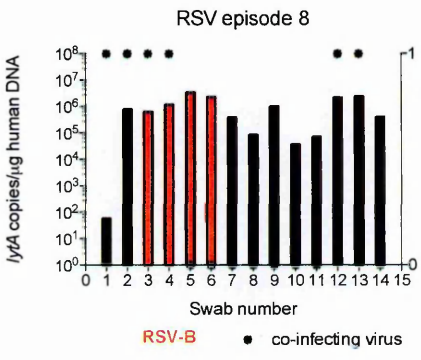

RSV episode 3
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

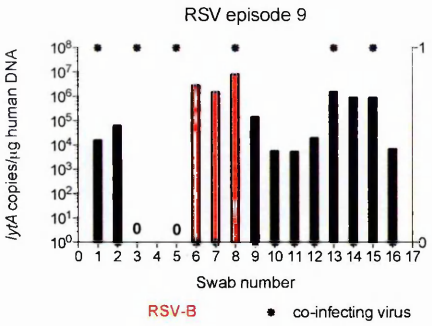

RSV episode 4
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

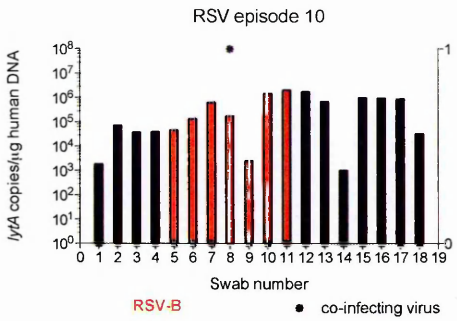

RSV episode 5
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

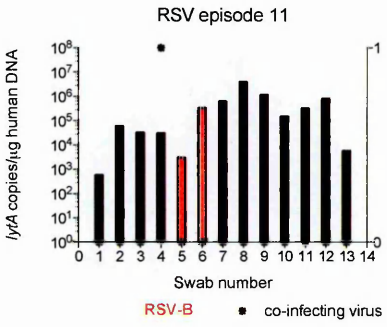

RSV episode 6
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

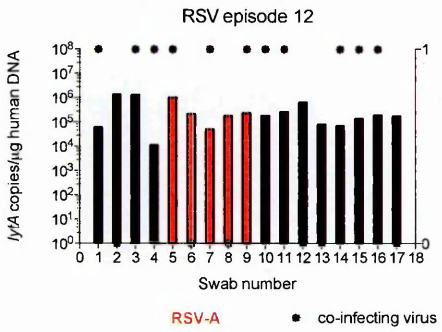

RSV episode 7
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

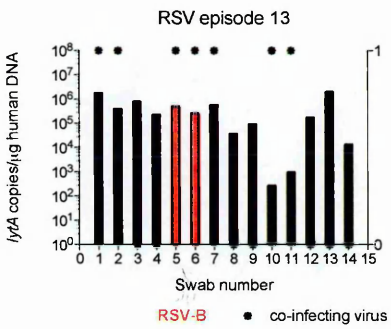

RSV episode 8
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

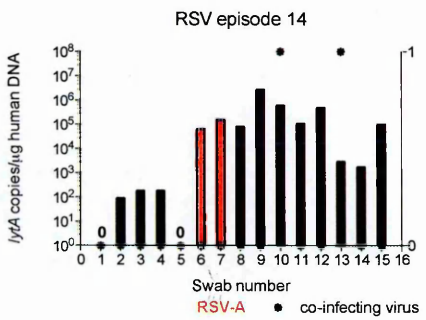

RSV episode 9
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

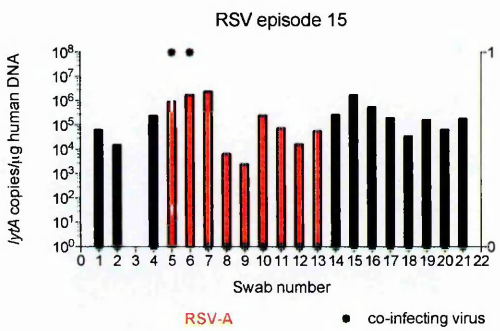

RSV episode 10
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

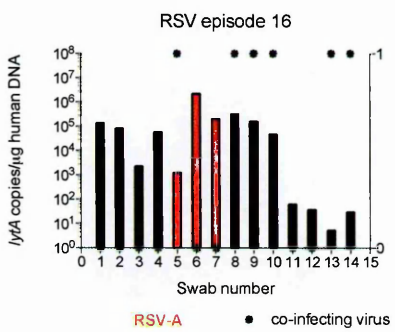

RSV episode 11
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

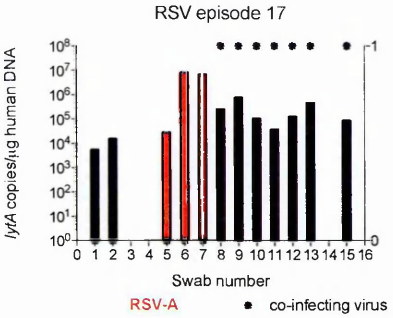

RSV episode 12
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

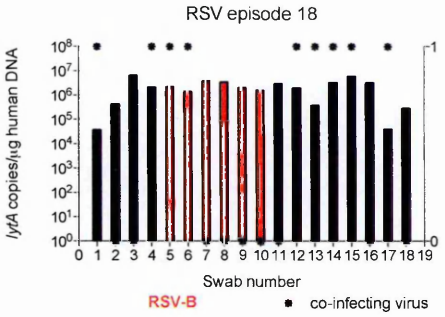

RSV episode 13
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

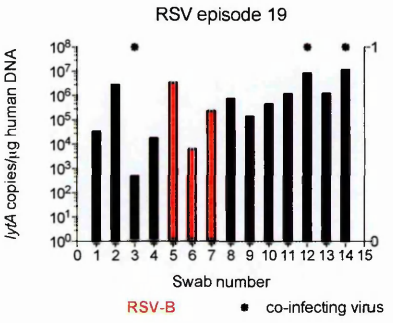

RSV episode 14
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

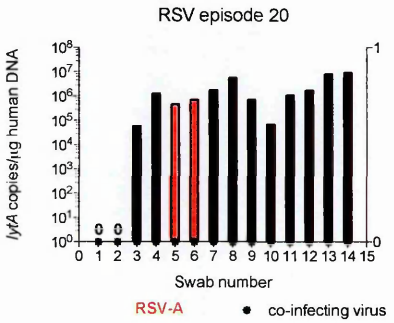

RSV episode 15
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

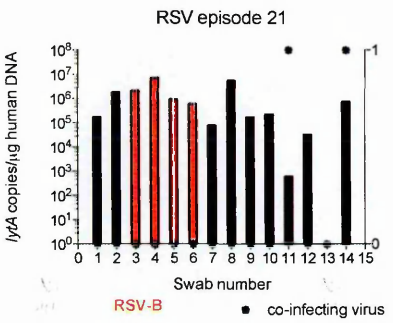

RSV episode 16
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

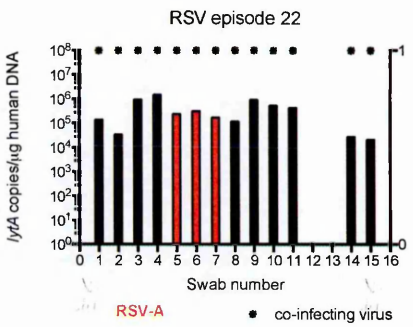

RSV episode 17
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

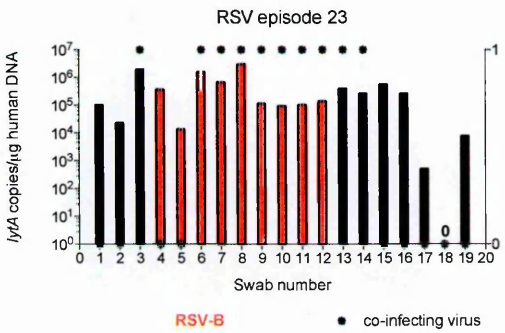

RSV episode 18
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

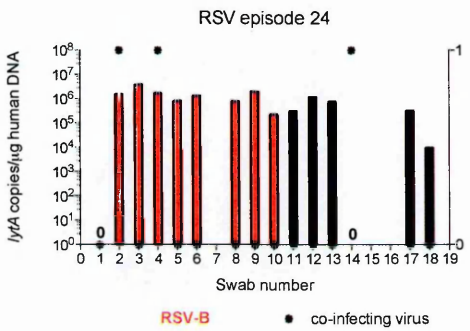

RSV episode 19
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

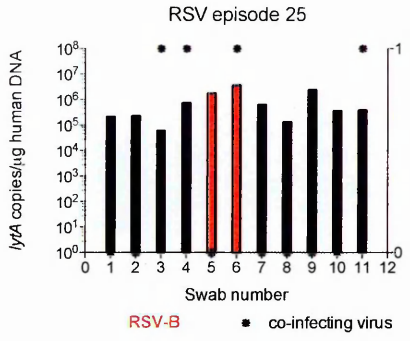

RSV episode 20
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

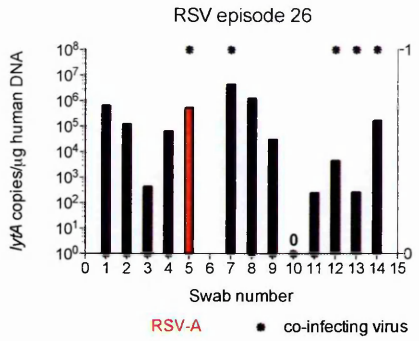

RSV episode 21
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

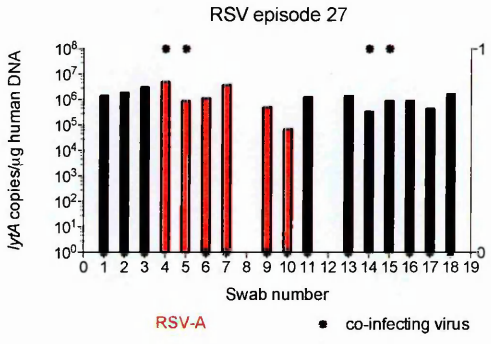

RSV episode 22
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

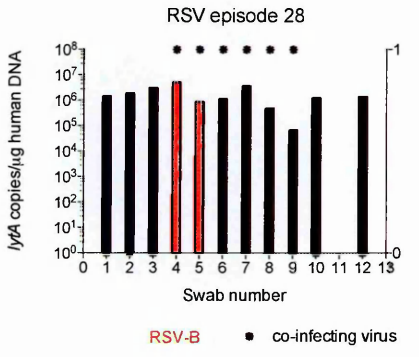

RSV episode 23
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

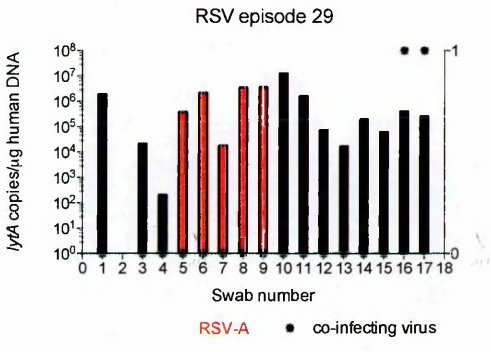

RSV episode 24
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

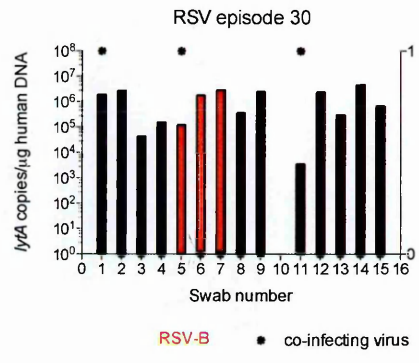

RSV episode 25
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

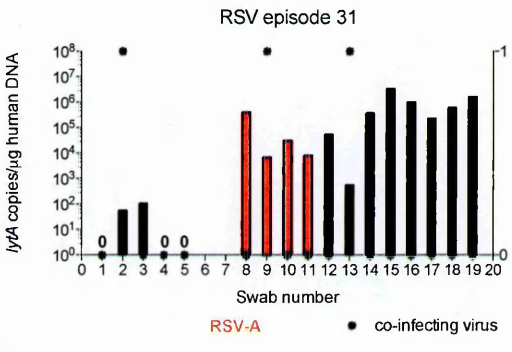

RSV episode 26
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

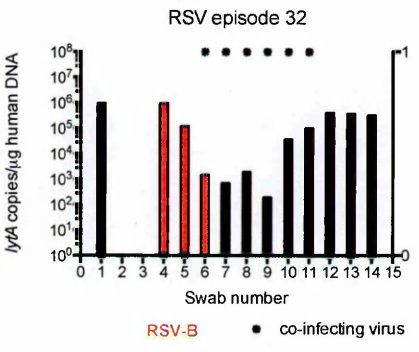

RSV episode 27
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

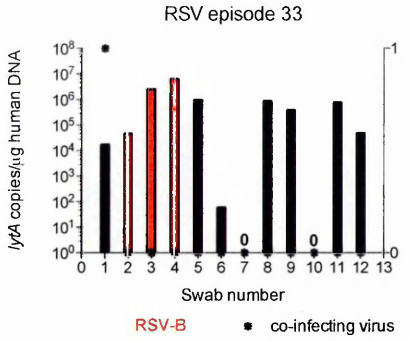

RSV episode 28
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

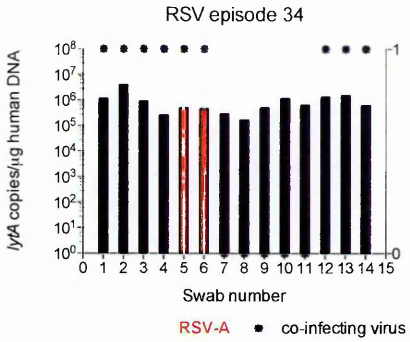

RSV episode 29
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

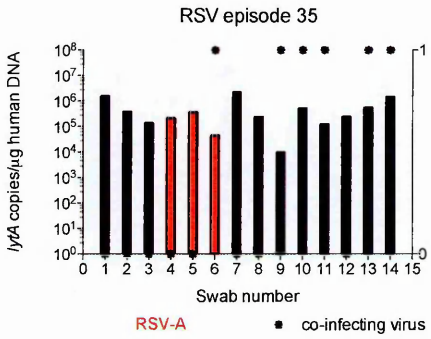

RSV episode 30
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

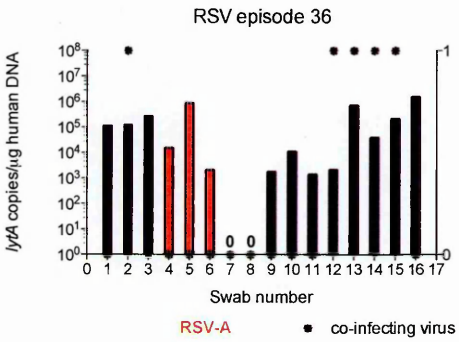

RSV episode 31
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

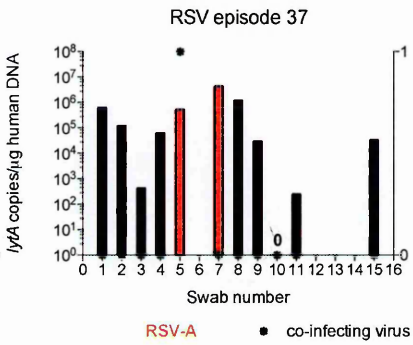

RSV episode 32
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

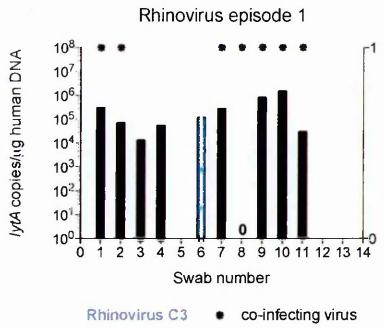

RSV episode 33
lytA copies/µg human DNA
Swab number
RSV-B
co-infecting virus

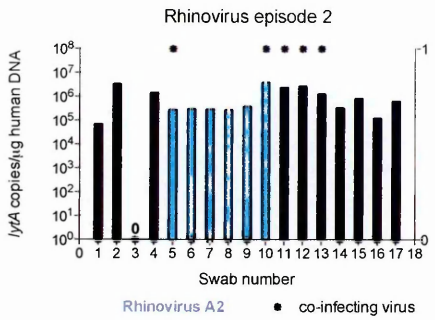

RSV episode 34
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

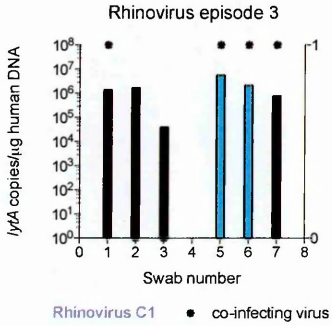

RSV episode 35
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

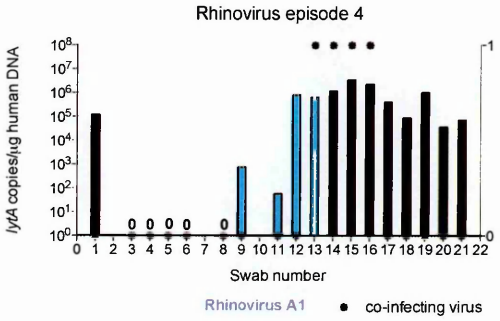

RSV episode 36
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

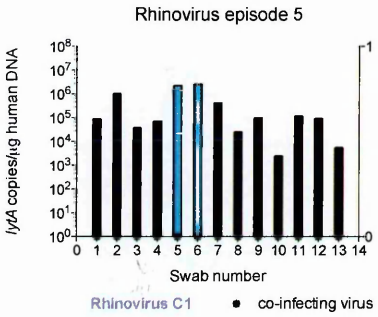

RSV episode 37
lytA copies/µg human DNA
Swab number
RSV-A
co-infecting virus

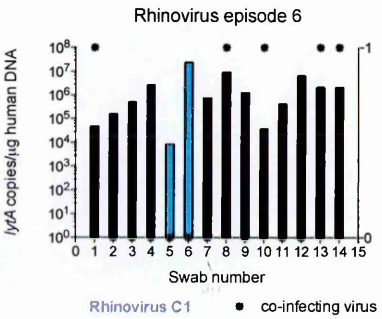

Rhinovirus episode 1
lytA copies/µg human DNA
Swab number
Rhinovirus C3
co-infecting virus

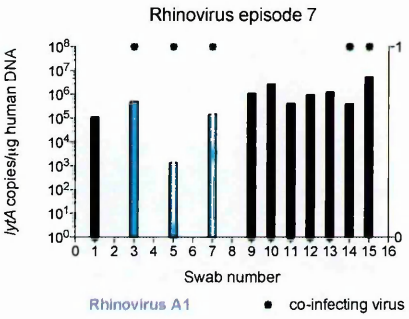

Rhinovirus episode 2
lytA copies/µg human DNA
Swab number
Rhinovirus A2
co-infecting virus

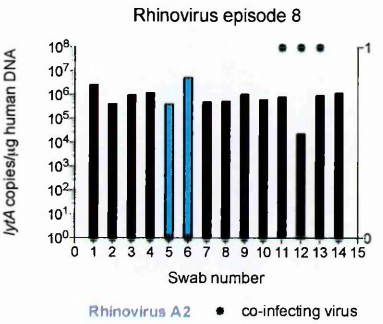

Rhinovirus episode 3
lytA copies/µg human DNA
Swab number
Rhinovirus C1
co-infecting virus

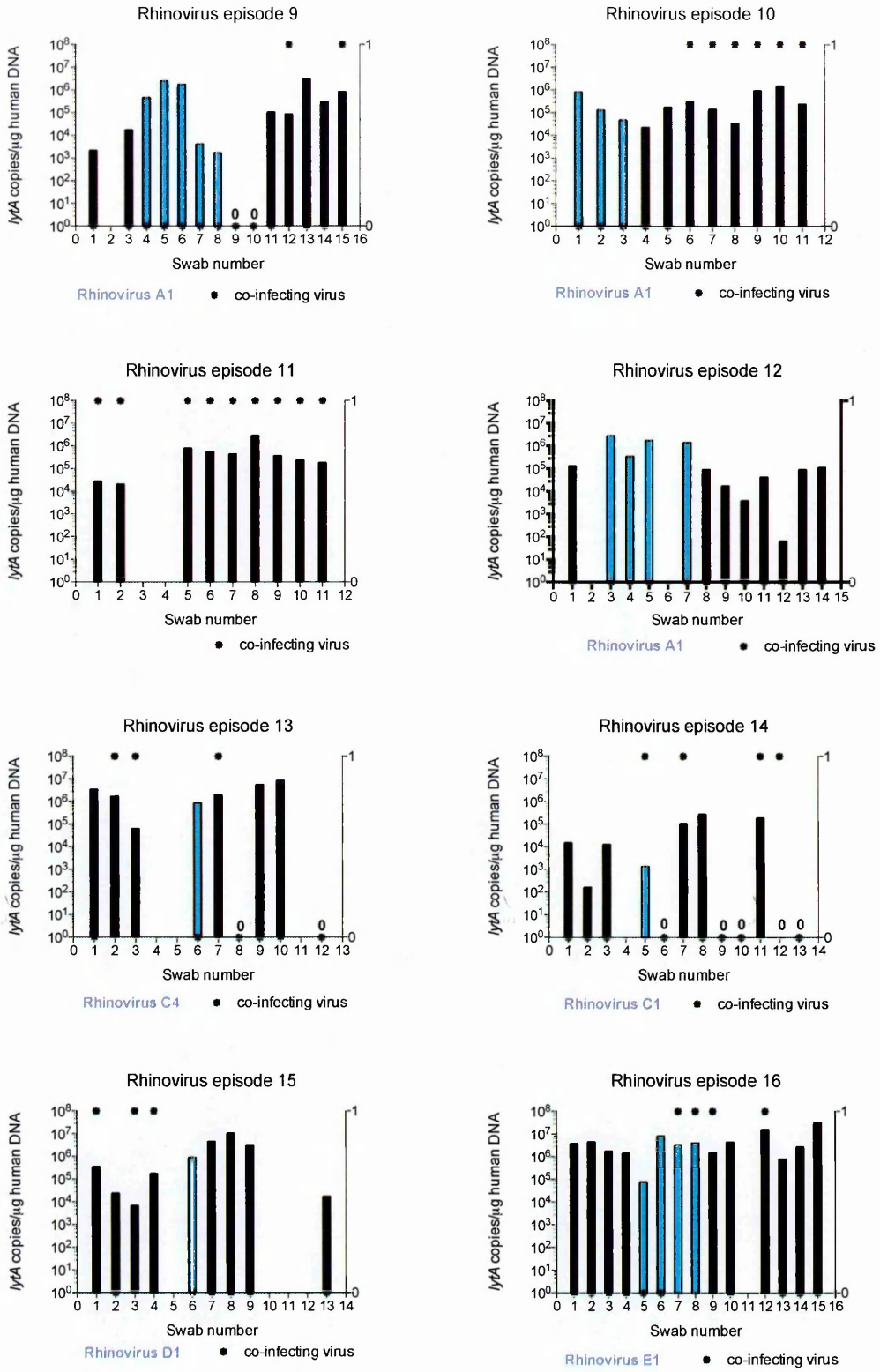

Rhinovirus episode 4
lytA copies/µg human DNA
Swab number
Rhinovirus A1
co-infecting virus


Rhinovirus episode 5
lytA copies/µg human DNA
Swab number
Rhinovirus C1
co-infecting virus


Rhinovirus episode 6
lytA copies/µg human DNA
Swab number
Rhinovirus C1
co-infecting virus


Rhinovirus episode 7
lytA copies/µg human DNA
Swab number
Rhinovirus A1
co-infecting virus

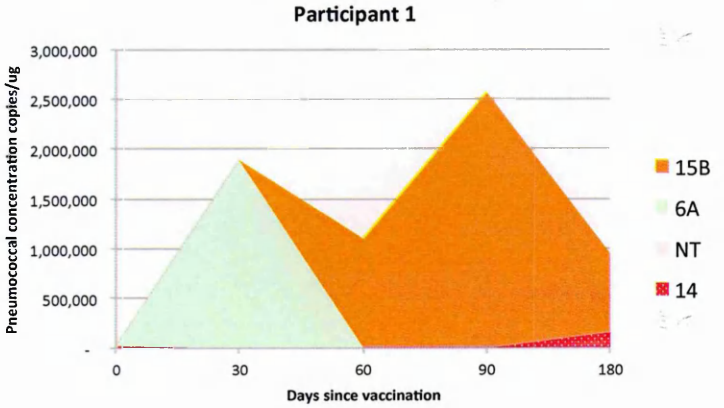

Rhinovirus episode 8
lytA copies/µg human DNA
Swab number
Rhinovirus A2
co-infecting virus



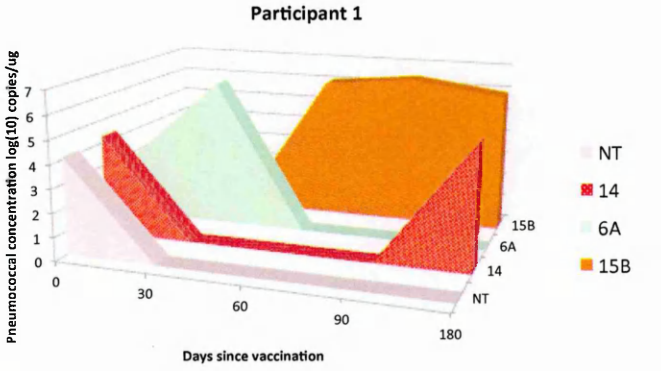
Appendix B: Nasopharyngeal pneumococcal concentration in *lytA* copies/µg of human DNA, before, during and after each studied episode of symptomatic upper respiratory tract infection (URTI) with RSV or rhinovirus in children <5 years old. The y-axis of each

graph is in $\log_{10}$ scale. The swab numbers on the x-axes are consecutive twice-weekly swabs. Where a swab was not collected or the sample was unavailable, the bar is missing. Where a swab was not missing but did not have detectable pneumococcus by *lytA* qPCR, it is marked zero. For RSV episodes, swabs collected during the episode are marked in red, for rhinovirus these are blue. Swabs two weeks before and four weeks after each viral episode are black. Bars marked with an asterisk indicate swabs that were co-infected with another respiratory virus by multiplex PCR.

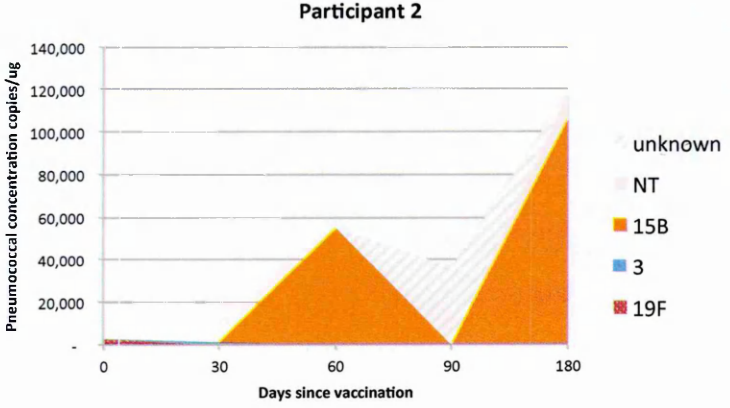
14.3 Appendix C: Pneumococcal concentration over 6 months post vaccination, by individual study participant



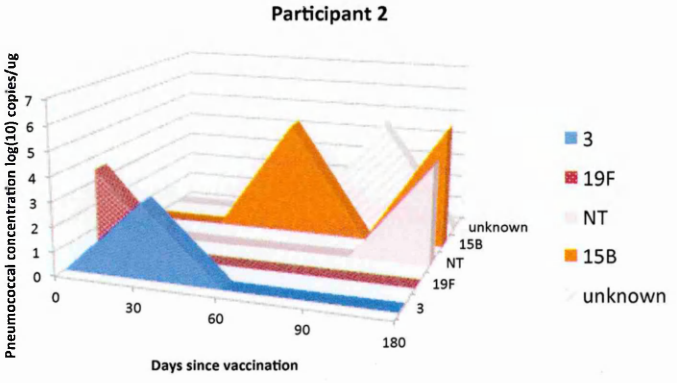
A



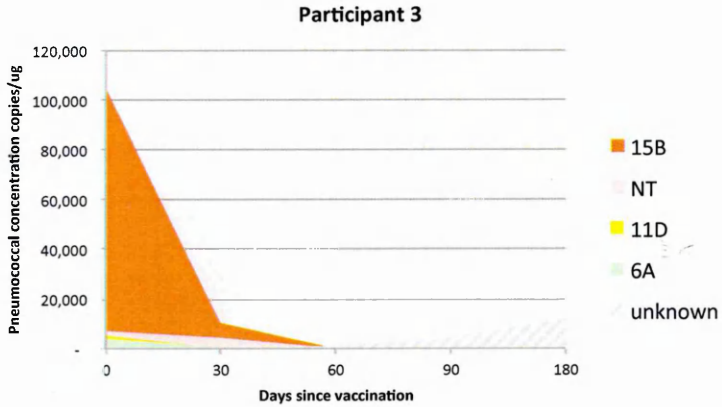
B



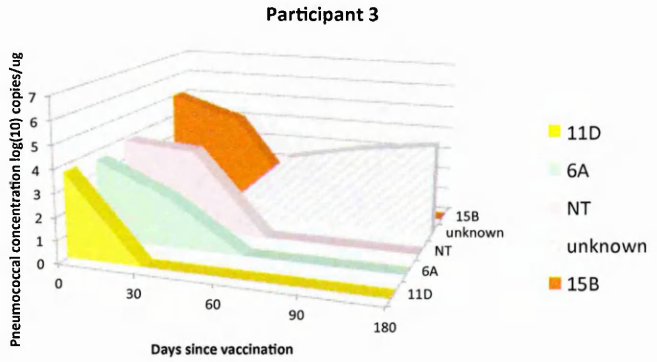
A



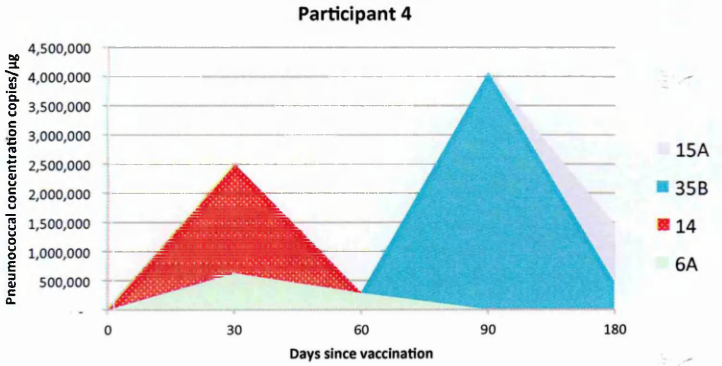
B


Participant 3
Pneumococcal concentration copies/ug
120,000
100,000
80,000
60,000
40,000
20,000
-
0
30
60
90
180
Days since vaccination
15B
NT
11D
6A
unknown

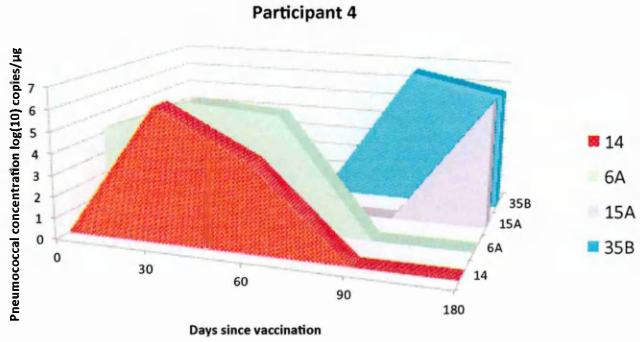
A


Participant 3
Pneumococcal concentration log(10) copies/ug
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15B
unknown
NT
6A
11D
11D
6A
NT
unknown
15B

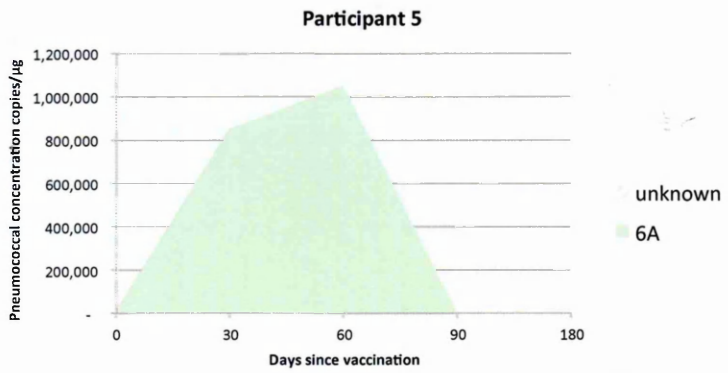
B


Participant 4
Pneumococcal concentration copies/µg
4,500,000
4,000,000
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
15A
35B
14
6A

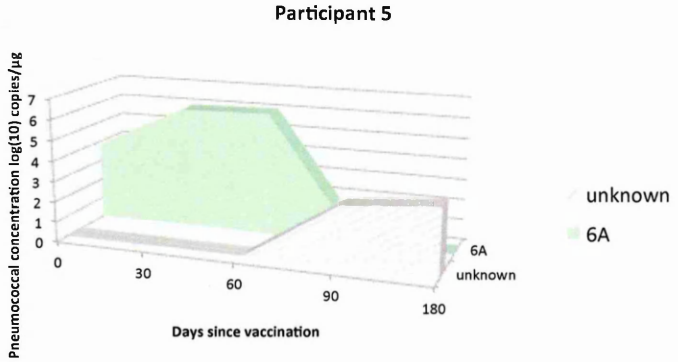
A


Participant 4
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
35B
15A
6A
14
14
6A
15A
35B

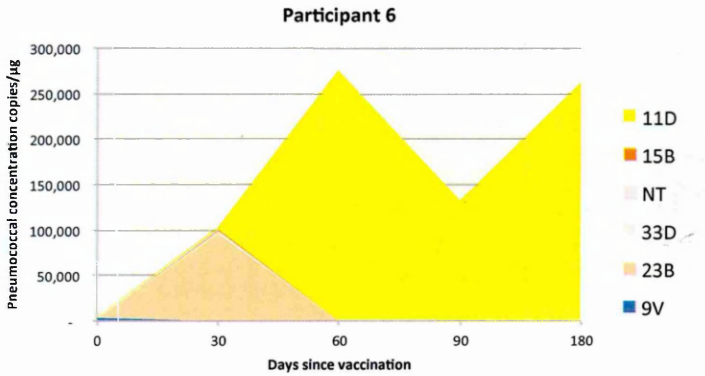
B


Participant 5
Pneumococcal concentration copies/µg
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
unknown
6A

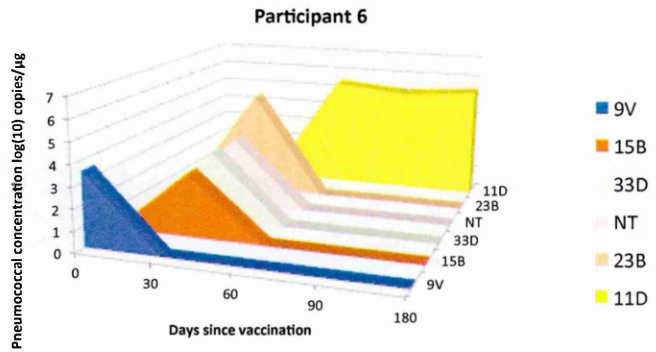
A


Participant 5
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
6A
unknown
unknown
6A

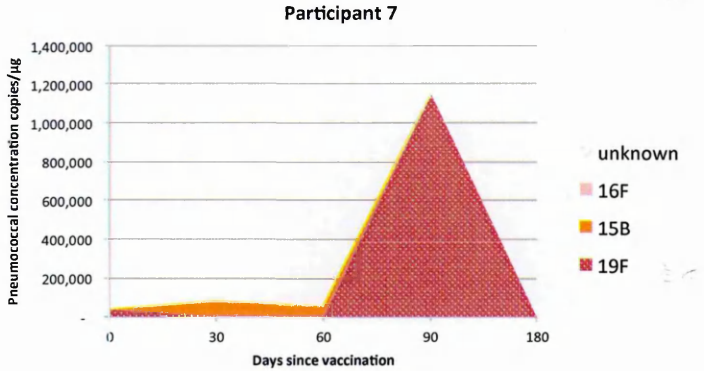
B


Participant 6
Pneumococcal concentration copies/µg
300,000
250,000
200,000
150,000
100,000
50,000
-
0
30
60
90
180
Days since vaccination
11D
15B
NT
33D
23B
9V

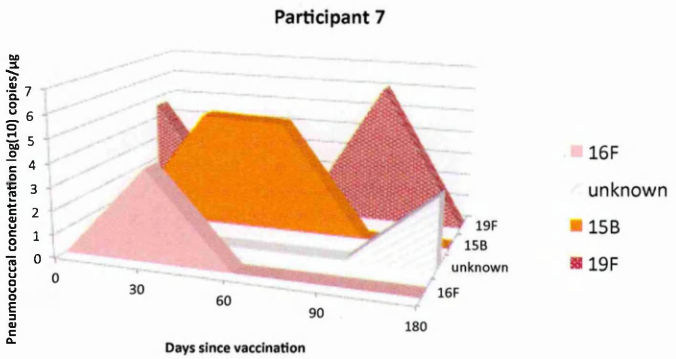
A


Participant 6
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
11D
23B
NT
33D
15B
9V
9V
15B
33D
NT
23B
11D

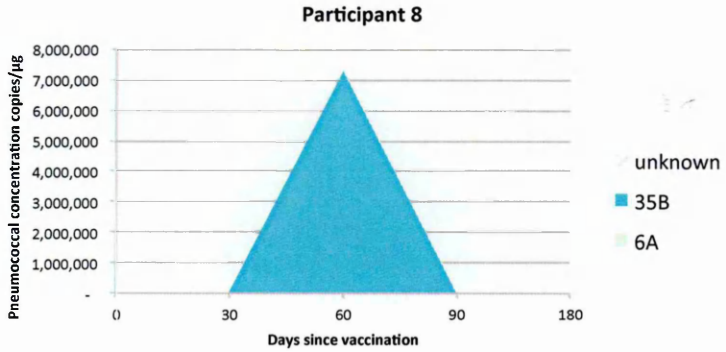
B


Participant 7
Pneumococcal concentration copies/µg
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
unknown
16F
15B
19F

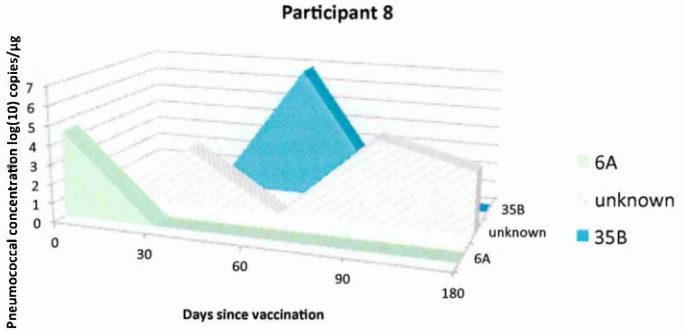
A


Participant 7
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
19F
15B
unknown
16F
16F
unknown
15B
19F

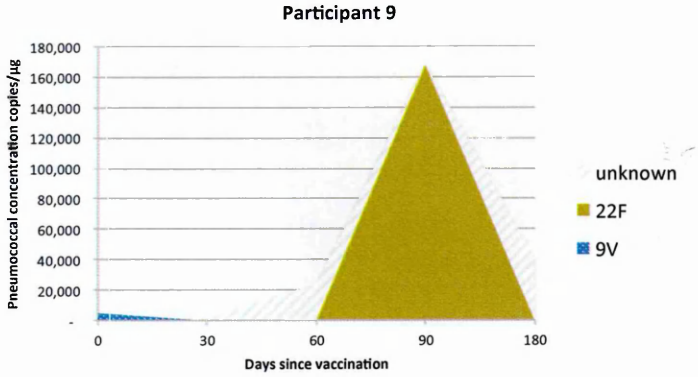
B


Participant 8
Pneumococcal concentration copies/µg
8,000,000
7,000,000
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
-
0
30
60
90
180
Days since vaccination
unknown
35B
6A

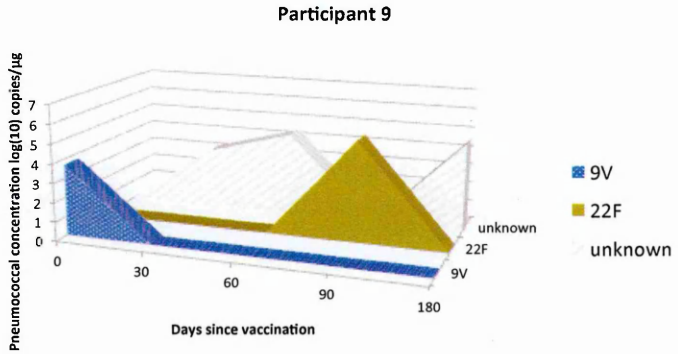
A


Participant 8
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
35B
unknown
6A
6A
unknown
35B

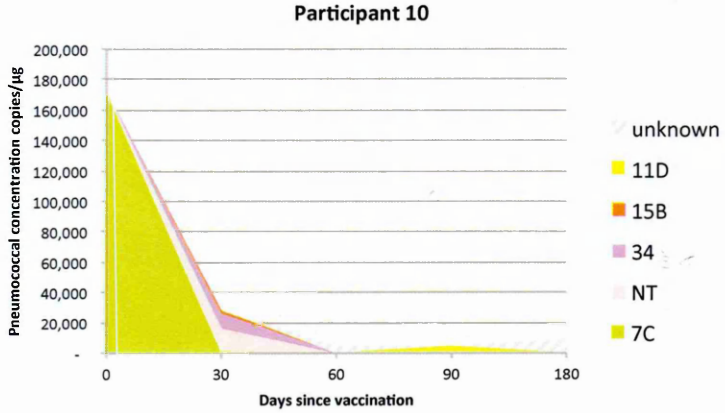
B


Participant 9
Pneumococcal concentration copies/µg
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
-
0
30
60
90
180
Days since vaccination
unknown
22F
9V

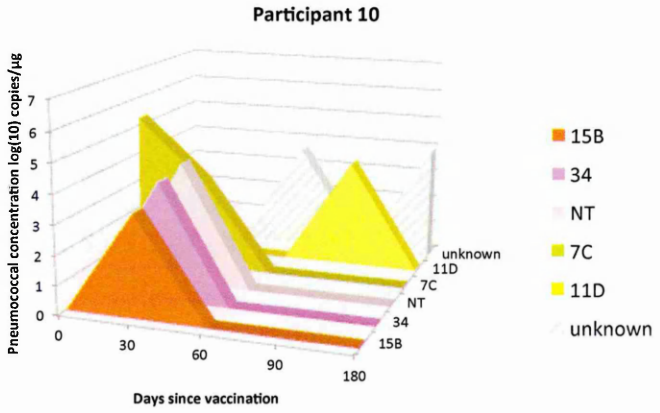
A


Participant 9
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
22F
9V
9V
22F
unknown

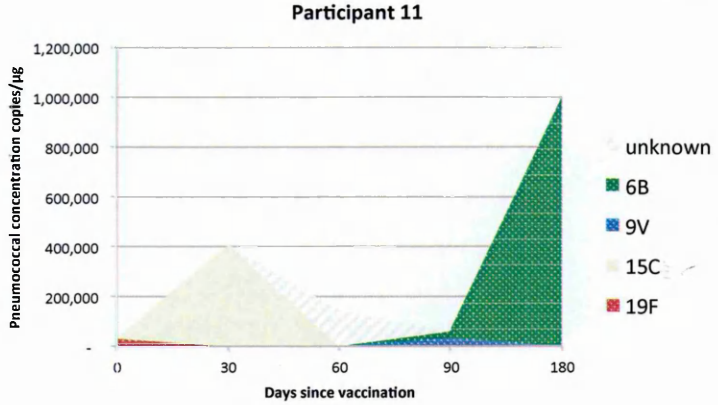
B


Participant 10
Pneumococcal concentration copies/µg
200,000
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
-
0
30
60
90
180
Days since vaccination
unknown
11D
15B
34
NT
7C

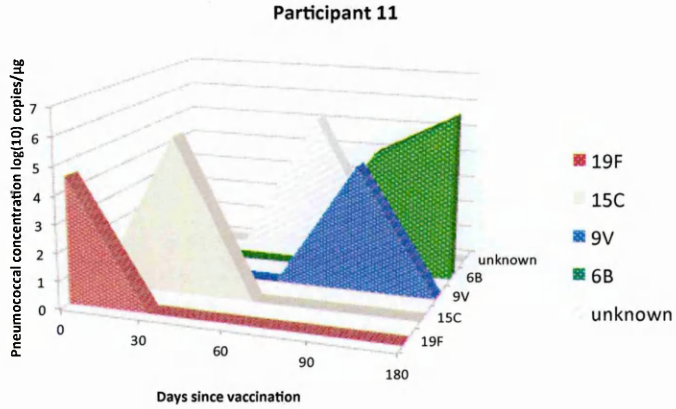
A


Participant 10
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
11D
7C
NT
34
15B
15B
34
NT
7C
11D
unknown

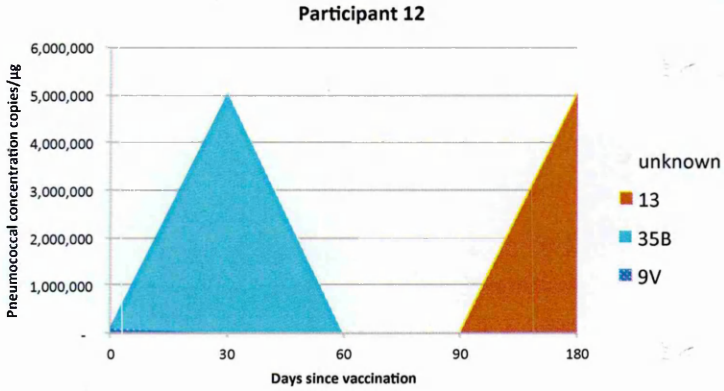
B


Participant 11
Pneumococcal concentration copies/µg
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
unknown
6B
9V
15C
19F

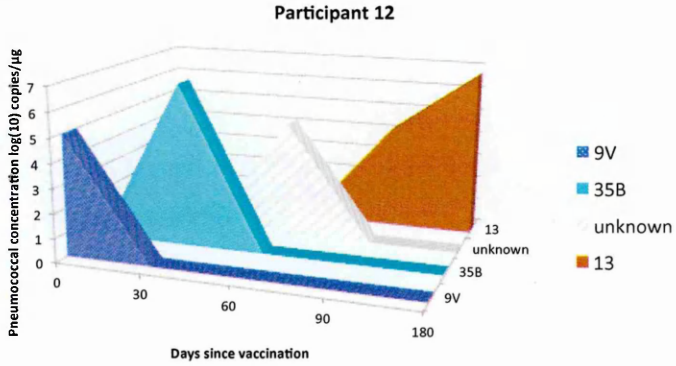
A


Participant 11
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
6B
9V
15C
19F
19F
15C
9V
6B
unknown

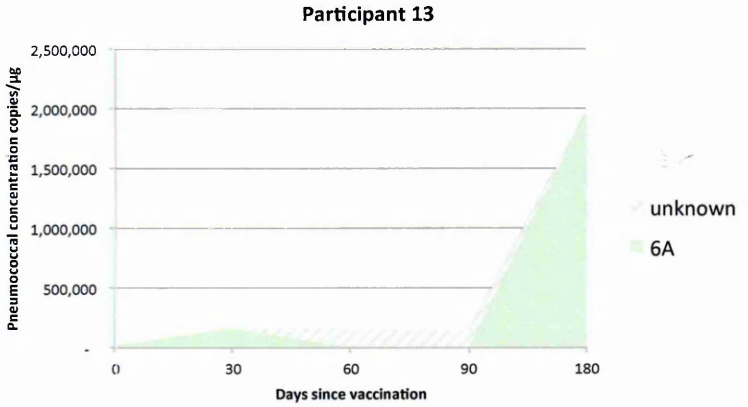
B


Participant 12
Pneumococcal concentration copies/µg
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
-
0
30
60
90
180
Days since vaccination
unknown
13
35B
9V

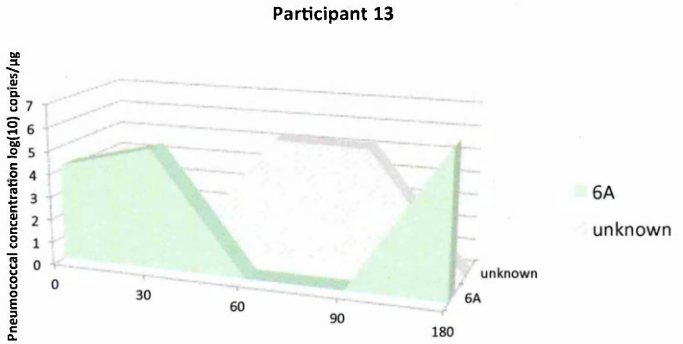
A


Participant 12
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
13
unknown
35B
9V
9V
35B
unknown
13

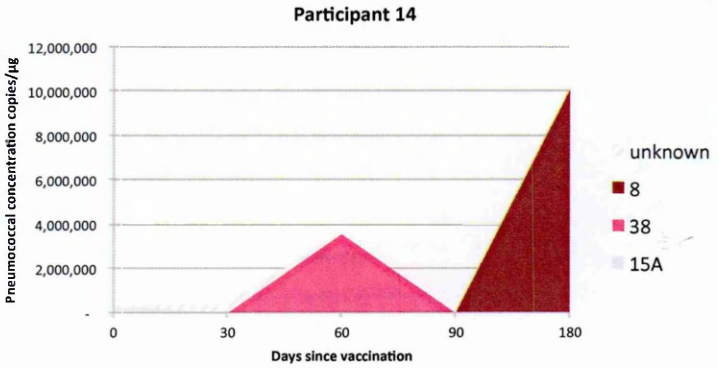
B


Participant 13
Pneumococcal concentration copies/µg
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
unknown
6A

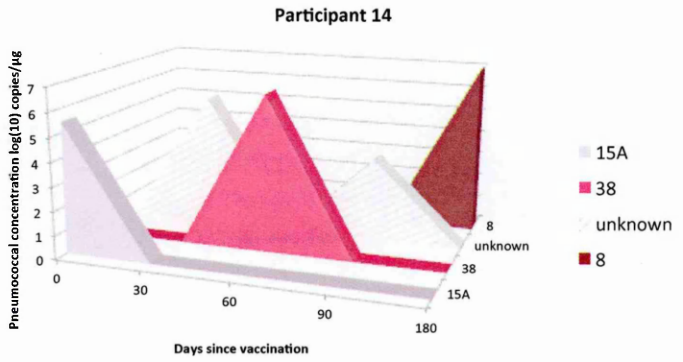
A


Participant 13
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
unknown
6A
6A
unknown

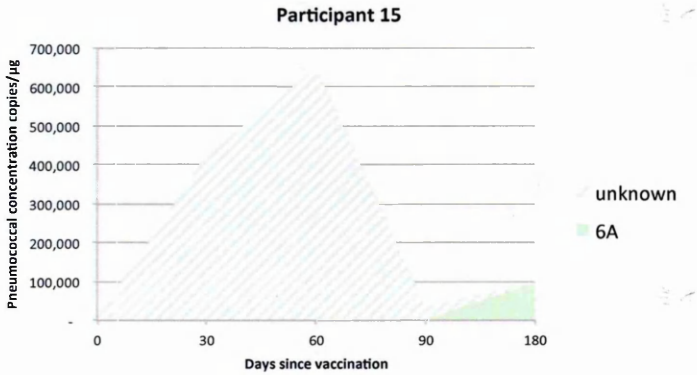
B


Participant 14
Pneumococcal concentration copies/µg
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
-
0
30
60
90
180
Days since vaccination
unknown
8
38
15A

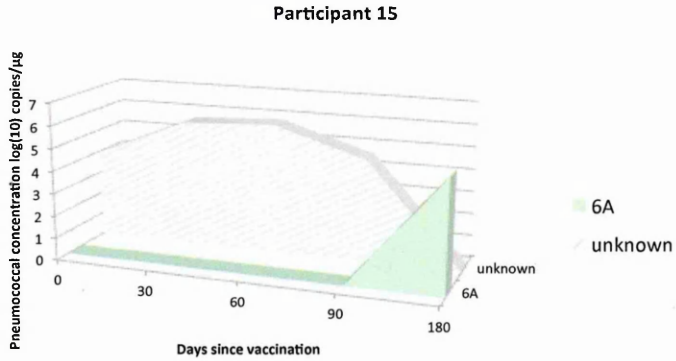
A


Participant 14
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
8
unknown
38
15A
15A
38
unknown
8

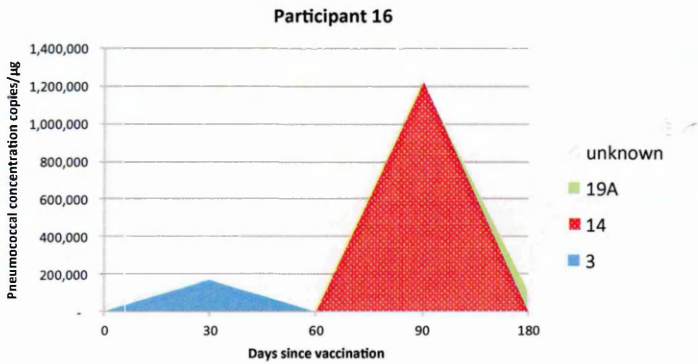
B


Participant 15
Pneumococcal concentration copies/µg
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
unknown
6A

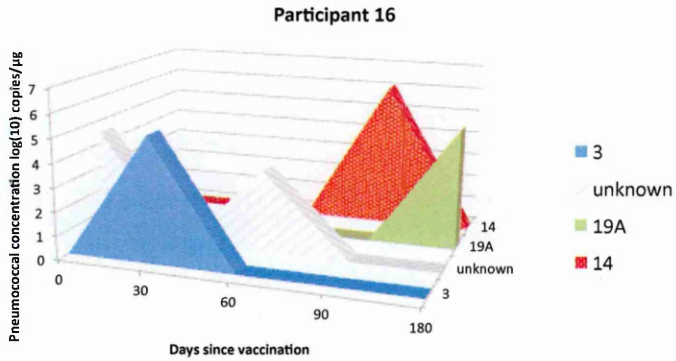
A


Participant 15
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
6A
6A
unknown

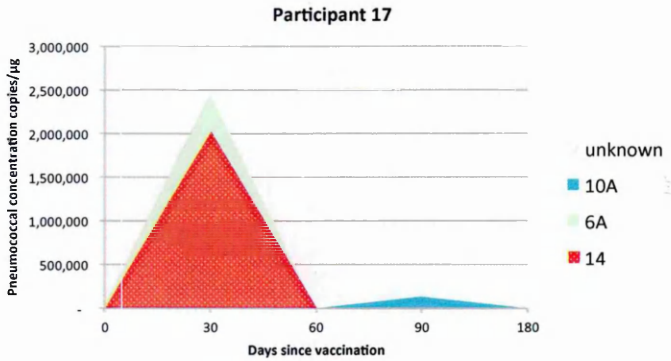
B


Participant 16
Pneumococcal concentration copies/µg
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
unknown
19A
14
3

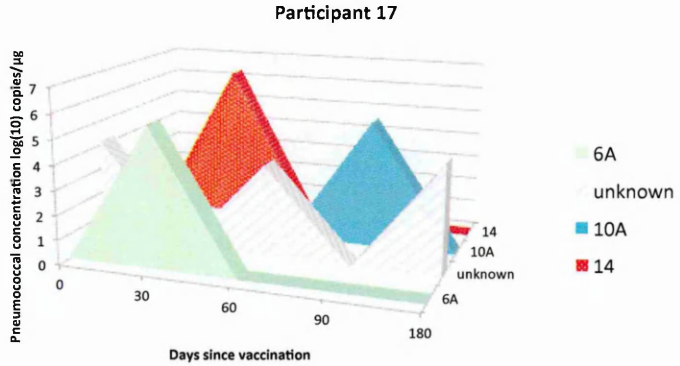
A


Participant 16
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
14
19A
unknown
3
3
unknown
19A
14

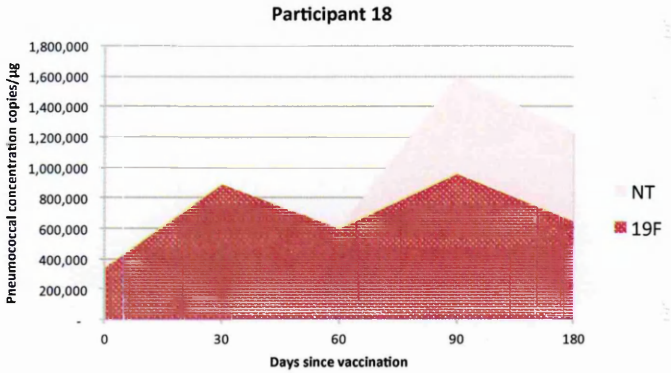
B


Participant 17
Pneumococcal concentration copies/µg
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
unknown
10A
6A
14

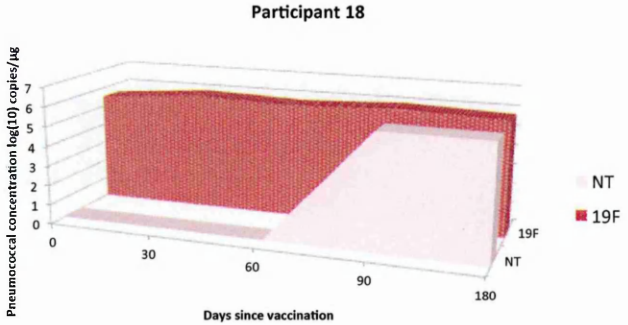
A


Participant 17
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
14
10A
unknown
6A
6A
unknown
10A
14

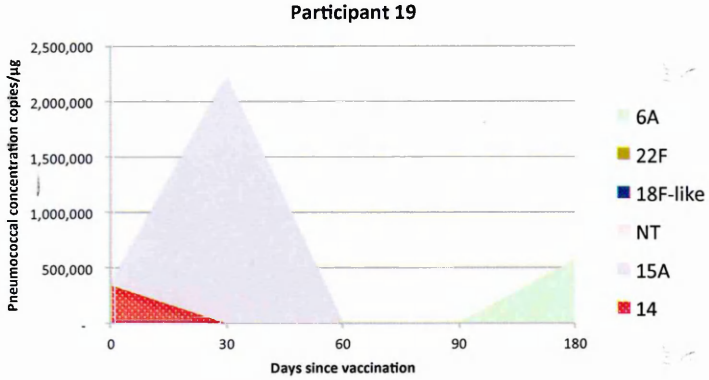
B


Participant 18
Pneumococcal concentration copies/µg
1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
NT
19F

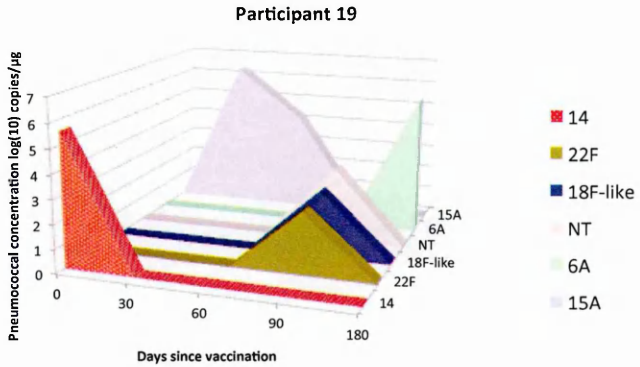
A


Participant 18
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
19F
NT
NT
19F

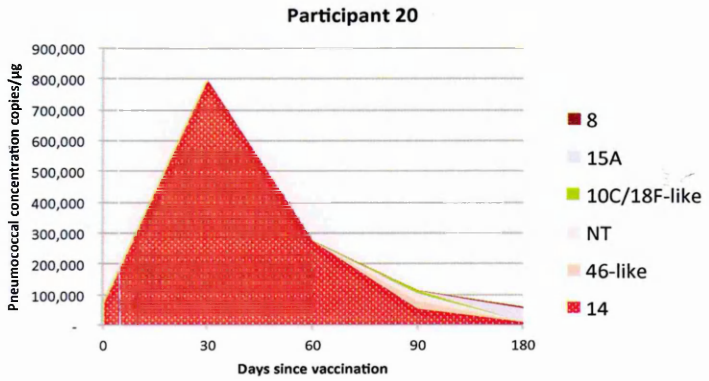
B


Participant 19
Pneumococcal concentration copies/µg
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
6A
22F
18F-like
NT
15A
14

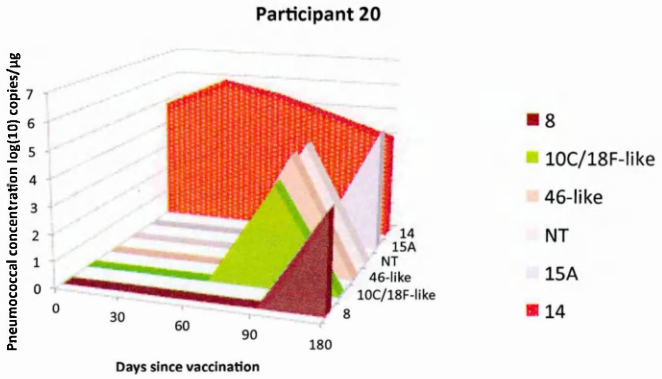
A


Participant 19
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15A
6A
NT
18F-like
22F
14
14
22F
18F-like
NT
6A
15A

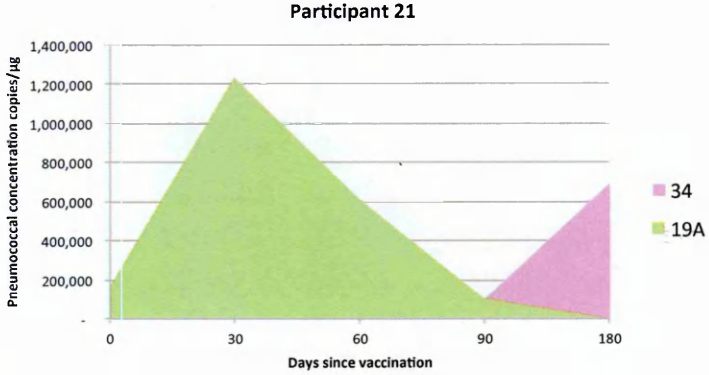
B


Participant 20
Pneumococcal concentration copies/µg
900,000
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
8
15A
10C/18F-like
NT
46-like
14

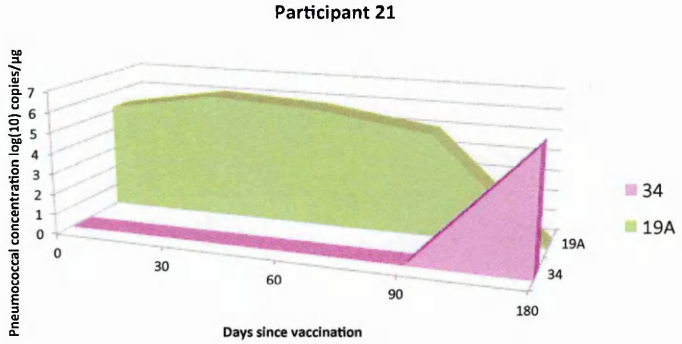
A


Participant 20
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
14
15A
NT
46-like
10C/18F-like
8
8
10C/18F-like
46-like
NT
15A
14

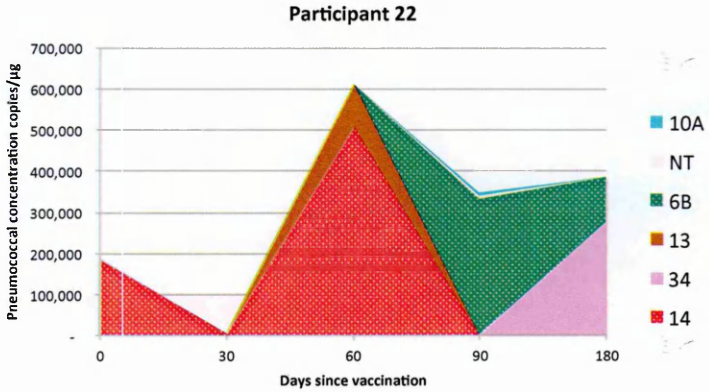
B


Participant 21
Pneumococcal concentration copies/µg
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
34
19A

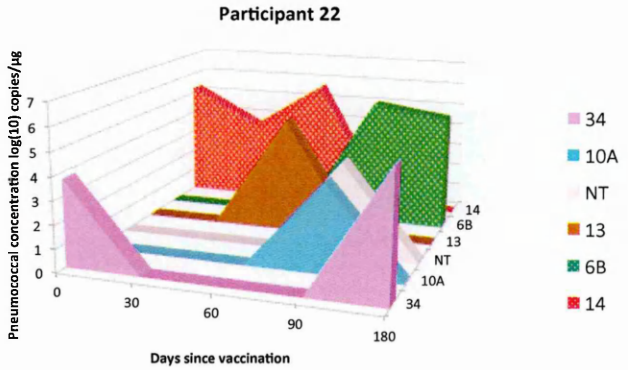
A


Participant 21
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
19A
34
34
19A

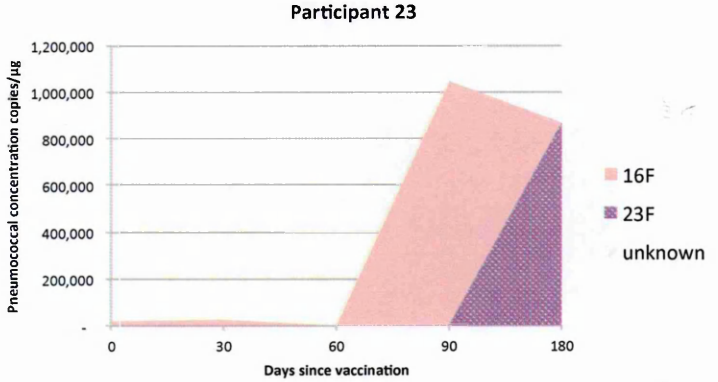
B


Participant 22
Pneumococcal concentration copies/µg
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
10A
NT
6B
13
34
14

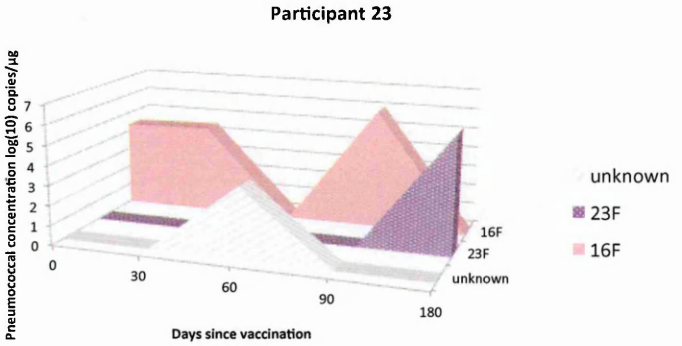
A


Participant 22
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
14
6B
13
NT
10A
34
34
10A
NT
13
6B
14

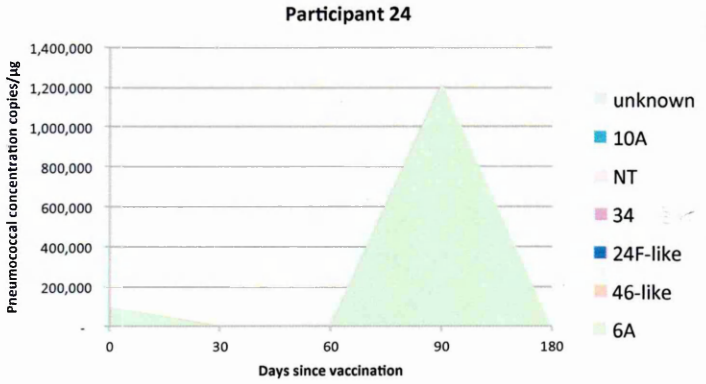
B


Participant 23
Pneumococcal concentration copies/µg
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
16F
23F
unknown

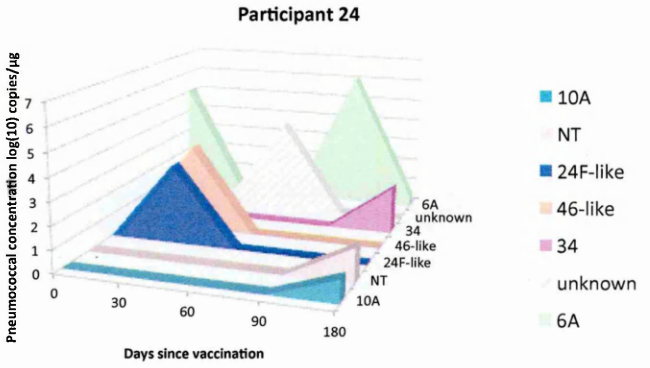
A


Participant 23
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
16F
23F
unknown
unknown
23F
16F

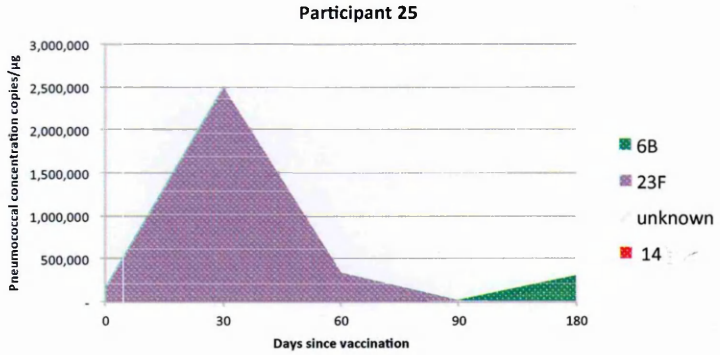
B


Participant 24
Pneumococcal concentration copies/µg
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
unknown
10A
NT
34
24F-like
46-like
6A

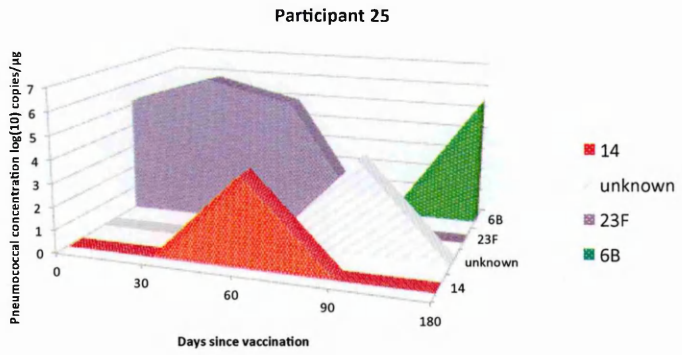
A


Participant 24
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
6A
unknown
34
46-like
24F-like
NT
10A
10A
NT
24F-like
46-like
34
unknown
6A

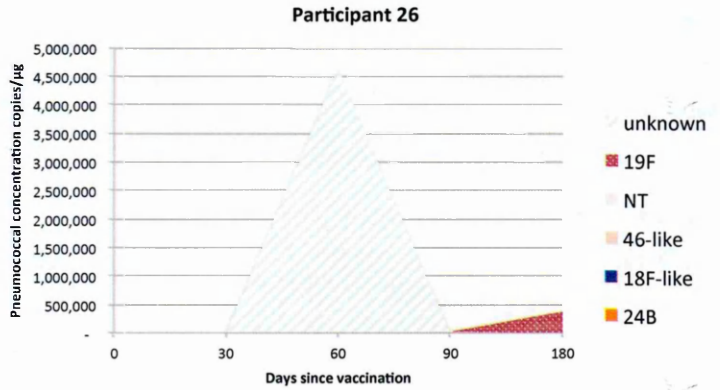
B


Participant 25
Pneumococcal concentration copies/µg
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
6B
23F
unknown
14

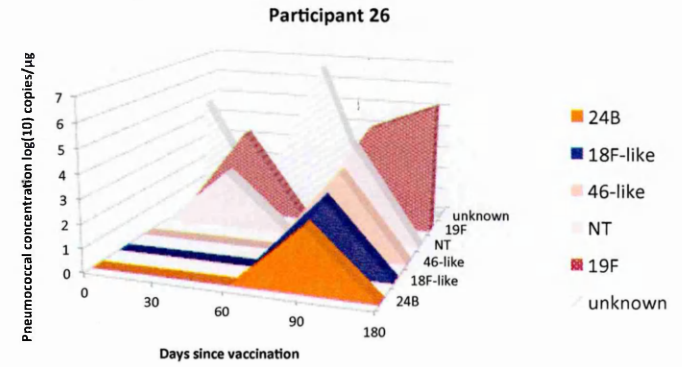
A


Participant 25
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
6B
23F
unknown
14
14
unknown
23F
6B

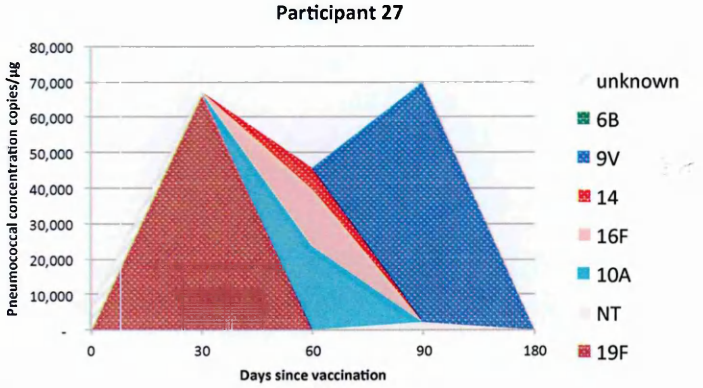
B


Participant 26
Pneumococcal concentration copies/µg
5,000,000
4,500,000
4,000,000
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
unknown
19F
NT
46-like
18F-like
24B

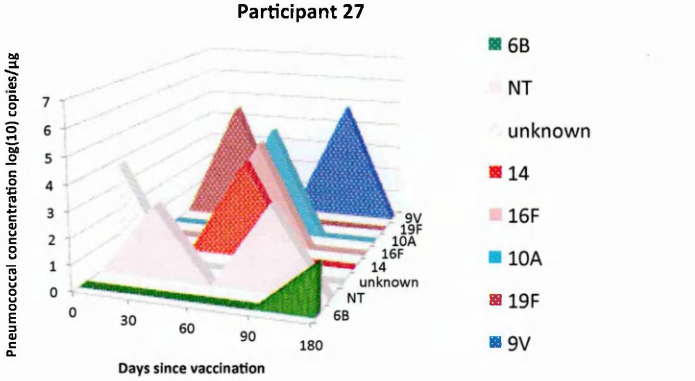
A


Participant 26
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
19F
NT
46-like
18F-like
24B
24B
18F-like
46-like
NT
19F
unknown

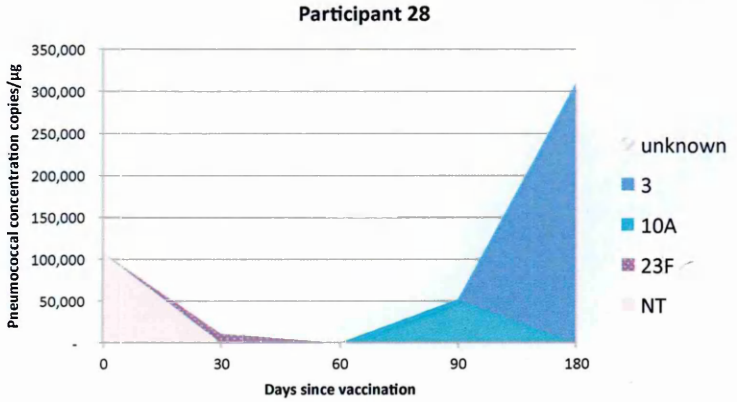
B


Participant 27
Pneumococcal concentration copies/µg
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
-
0
30
60
90
180
Days since vaccination
unknown
6B
9V
14
16F
10A
NT
19F

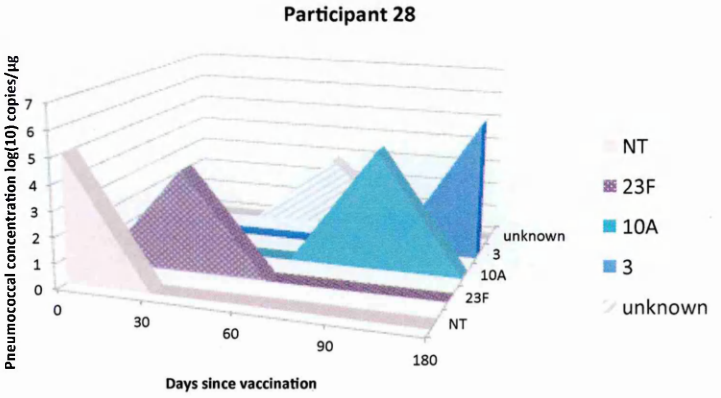
A


Participant 27
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
9V
19F
10A
16F
14
unknown
NT
6B
6B
NT
unknown
14
16F
10A
19F
9V

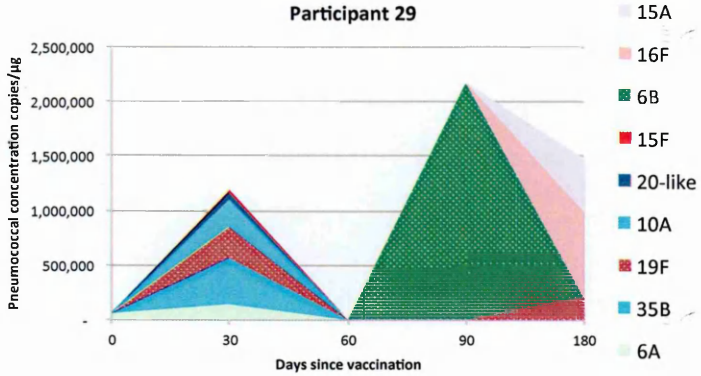
B


Participant 28
Pneumococcal concentration copies/µg
350,000
300,000
250,000
200,000
150,000
100,000
50,000
-
0
30
60
90
180
Days since vaccination
unknown
3
10A
23F
NT

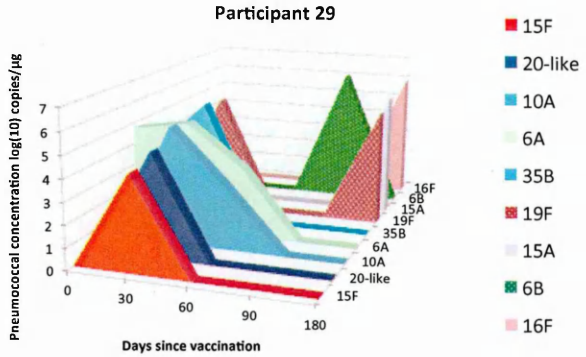
A


Participant 28
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
3
10A
23F
NT
NT
23F
10A
3
unknown

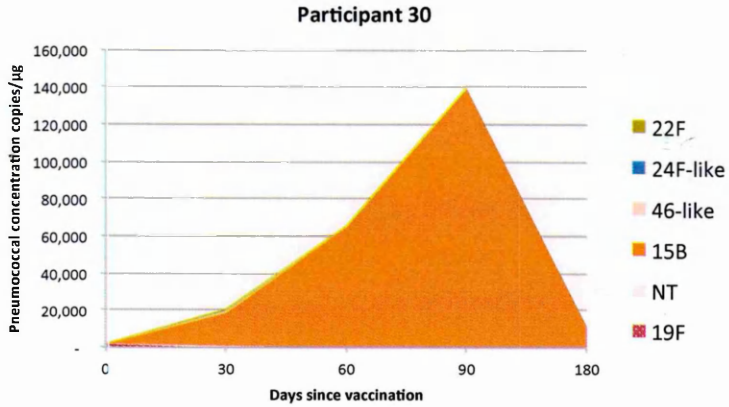
B


Participant 29
Pneumococcal concentration copies/µg
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
15A
16F
6B
15F
20-like
10A
19F
35B
6A

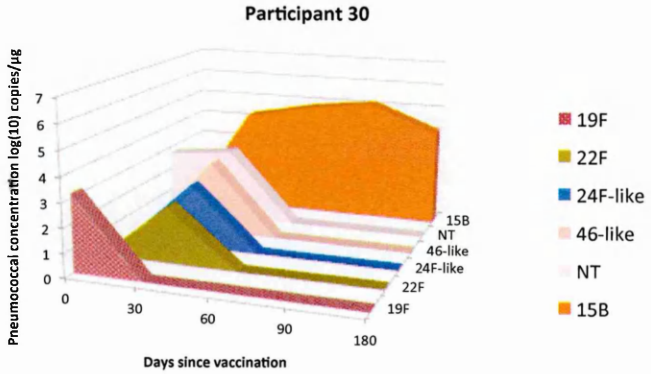
A


Participant 29
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
16F
6B
15A
19F
35B
6A
10A
20-like
15F
15F
20-like
10A
6A
35B
19F
15A
6B
16F

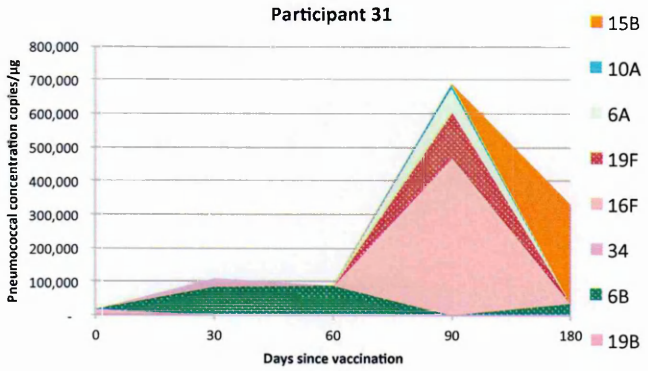
B


Participant 30
Pneumococcal concentration copies/µg
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
-
0
30
60
90
180
Days since vaccination
22F
24F-like
46-like
15B
NT
19F

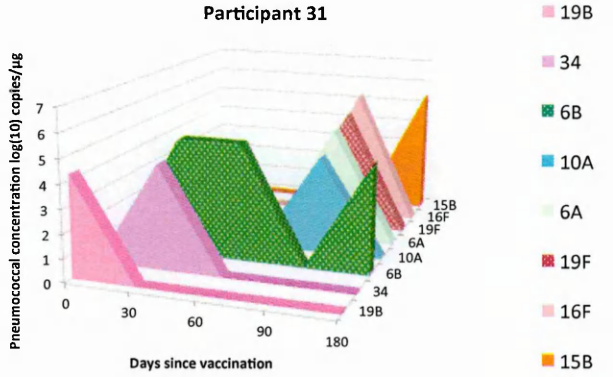
A


Participant 30
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15B
NT
46-like
24F-like
22F
19F
19F
22F
24F-like
46-like
NT
15B

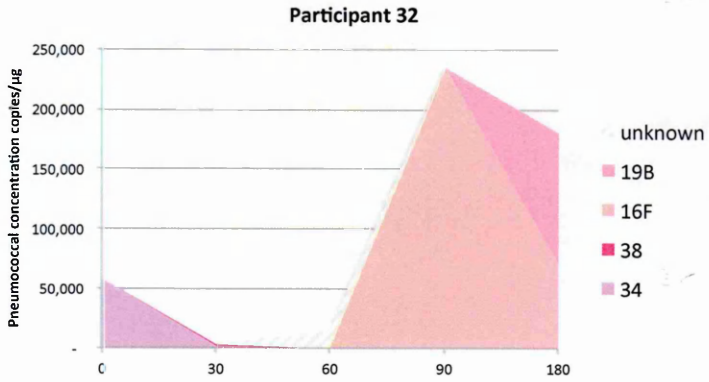
B


Participant 31
Pneumococcal concentration copies/µg
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
15B
10A
6A
19F
16F
34
6B
19B

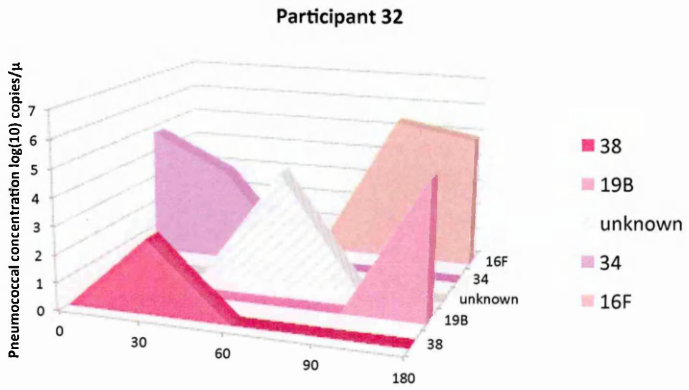
A


Participant 31
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15B
16F
19F
6A
10A
6B
34
19B
19B
34
6B
10A
6A
19F
16F
15B

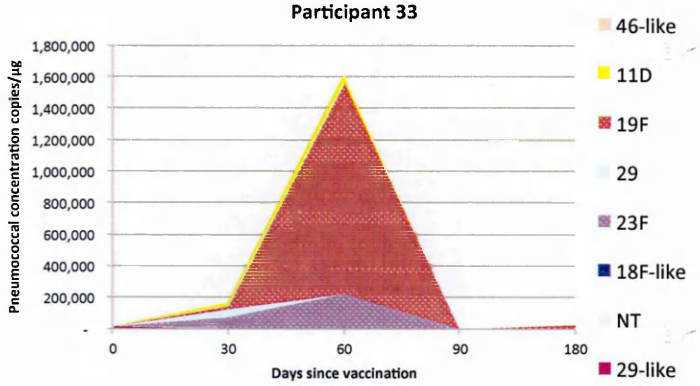
B


Participant 32
Pneumococcal concentration copies/µg
250,000
200,000
150,000
100,000
50,000
-
0
30
60
90
180
unknown
19B
16F
38
34

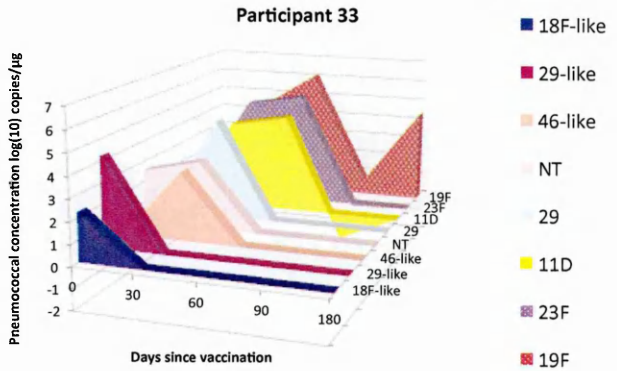
A


Participant 32
Pneumococcal concentration log(10) copies/µ
7
6
5
4
3
2
1
0
0
30
60
90
180
16F
34
unknown
19B
38
38
19B
unknown
34
16F

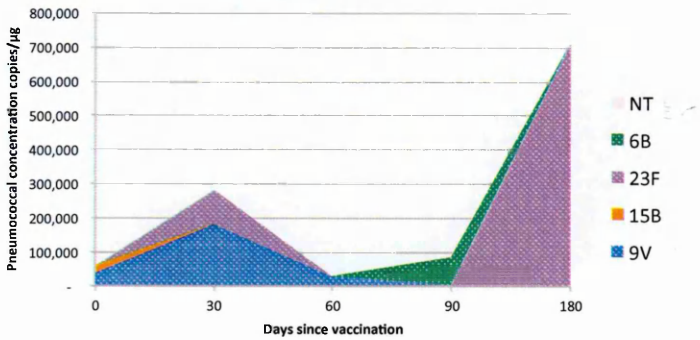
B


Participant 33
Pneumococcal concentration copies/µg
1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
46-like
11D
19F
29
23F
18F-like
NT
29-like

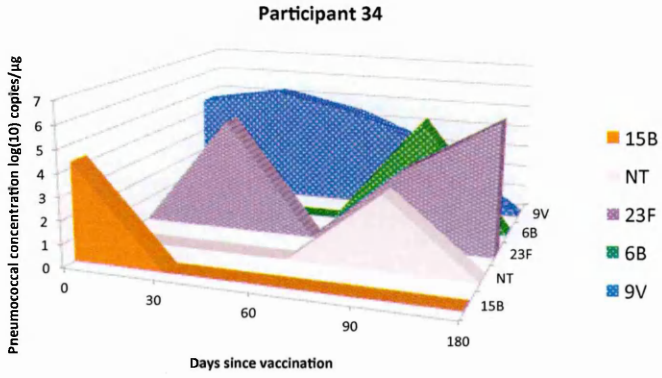
A


Participant 33
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
-1
-2
0
30
60
90
180
Days since vaccination
19F
23F
11D
29
NT
46-like
29-like
18F-like
18F-like
29-like
46-like
NT
29
11D
23F
19F

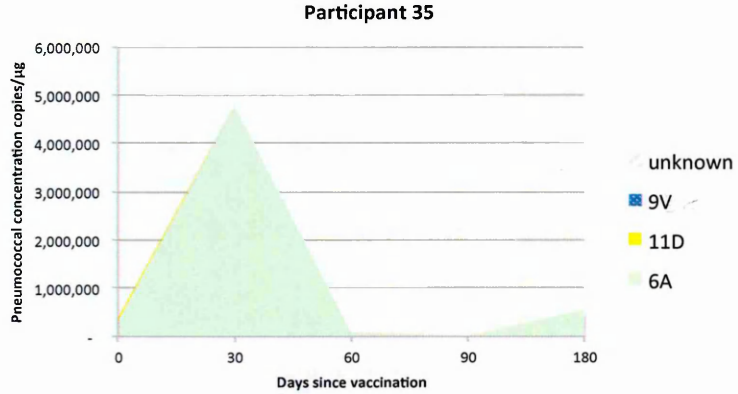
B


Participant 34
Pneumococcal concentration copies/µg
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
NT
6B
23F
15B
9V

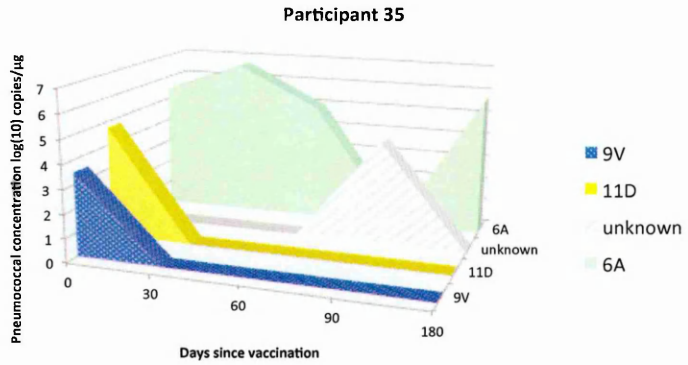
A


Participant 34
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
9V
6B
23F
NT
15B
15B
NT
23F
6B
9V

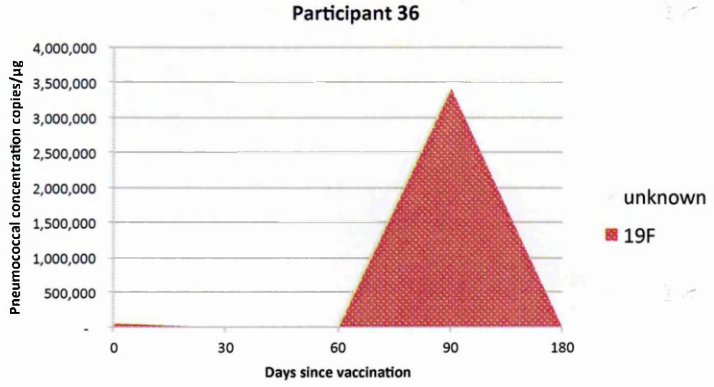
B


Participant 35
Pneumococcal concentration copies/µg
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
-
0
30
60
90
180
Days since vaccination
unknown
9V
11D
6A

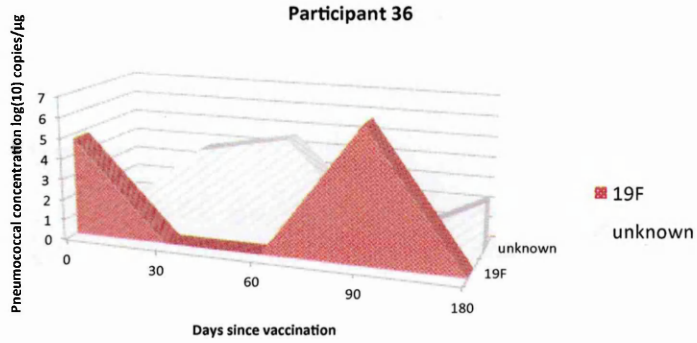
A


Participant 35
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
6A
unknown
11D
9V
9V
11D
unknown
6A

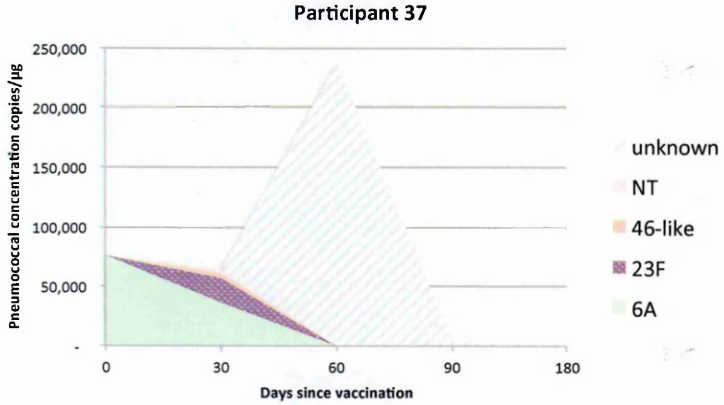
B


Participant 36
Pneumococcal concentration copies/µg
4,000,000
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
unknown
19F

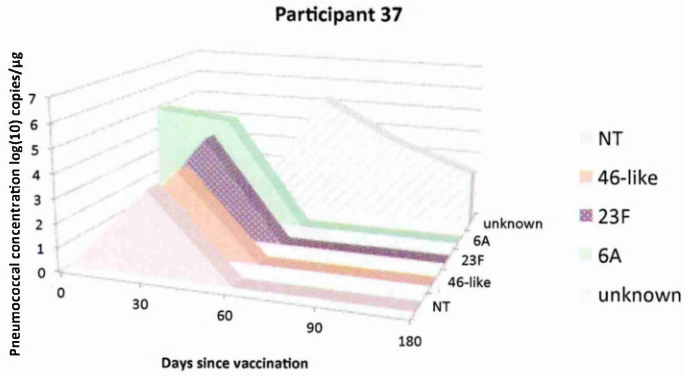
A


Participant 36
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
19F
19F
unknown

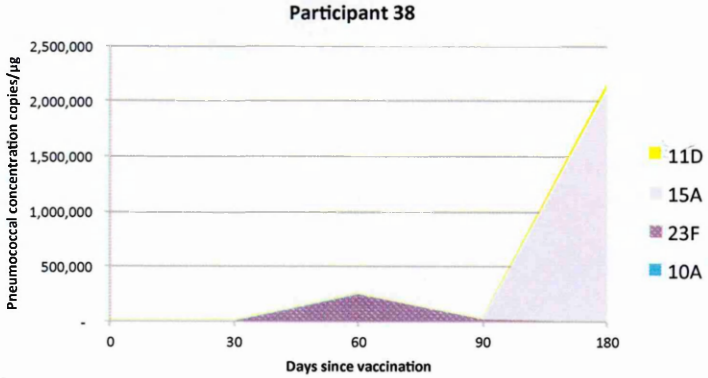
B


Participant 37
Pneumococcal concentration copies/µg
250,000
200,000
150,000
100,000
50,000
-
0
30
60
90
180
Days since vaccination
unknown
NT
46-like
23F
6A

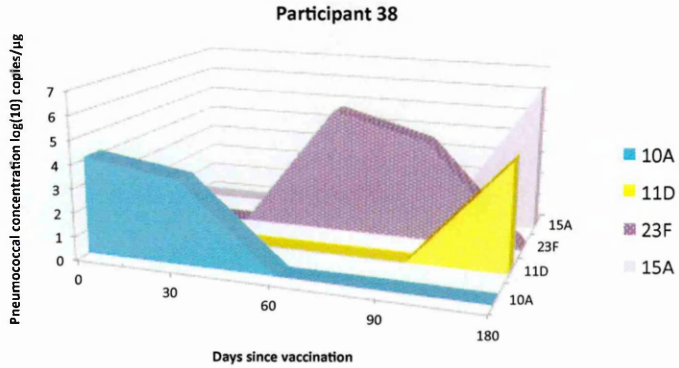
A


Participant 37
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
6A
23F
46-like
NT
NT
46-like
23F
6A
unknown

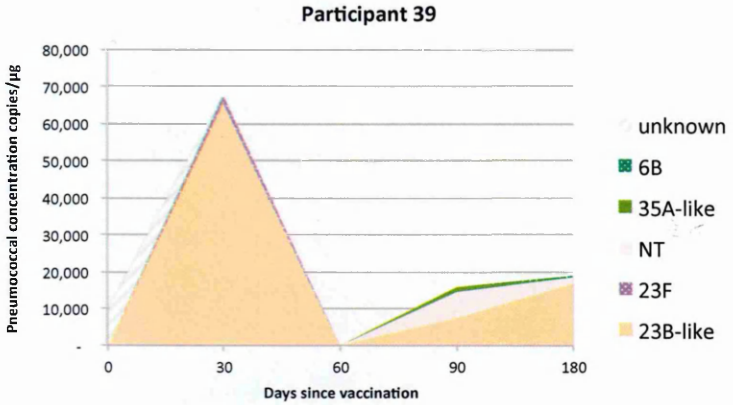
B


Participant 38
Pneumococcal concentration copies/µg
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
11D
15A
23F
10A

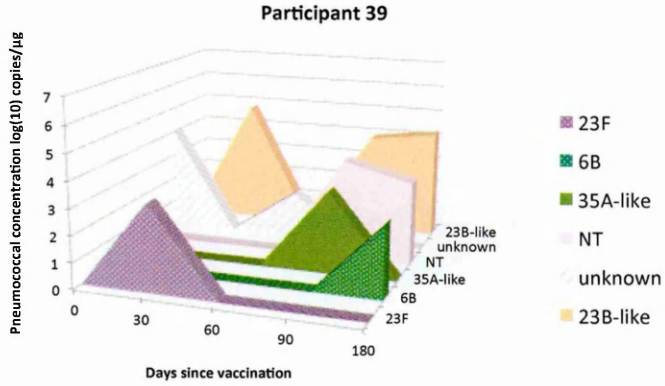
A


Participant 38
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15A
23F
11D
10A
10A
11D
23F
15A

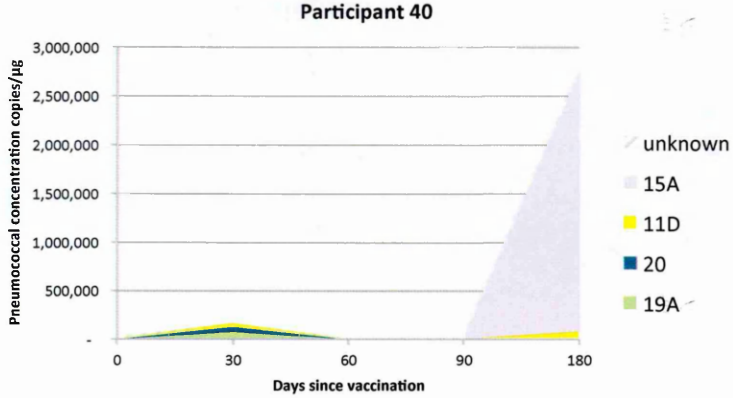
B


Participant 39
Pneumococcal concentration copies/µg
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
-
0
30
60
90
180
Days since vaccination
unknown
6B
35A-like
NT
23F
23B-like

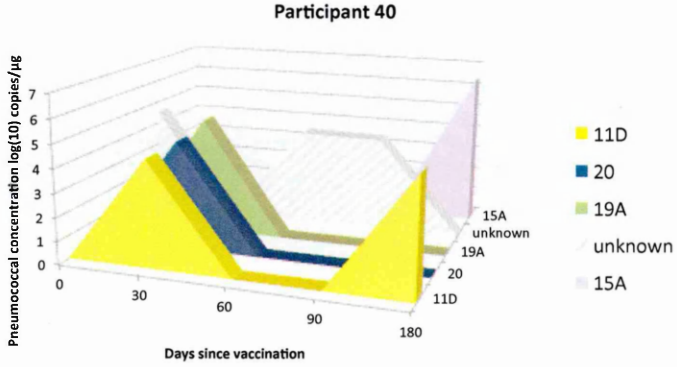
A


Participant 39
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
23B-like
unknown
NT
35A-like
6B
23F
23F
6B
35A-like
NT
unknown
23B-like

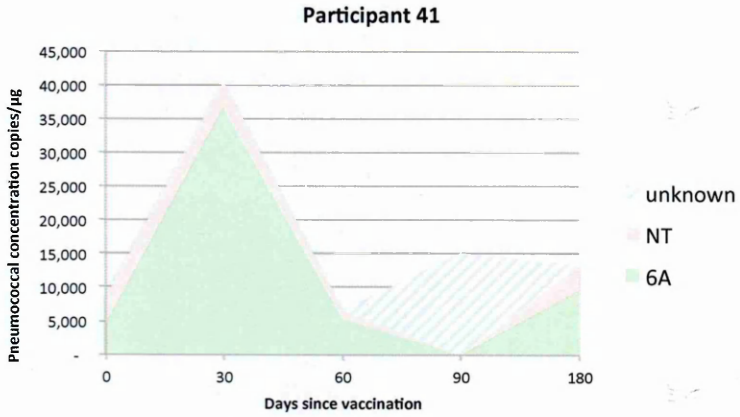
B


Participant 40
Pneumococcal concentration copies/µg
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
unknown
15A
11D
20
19A

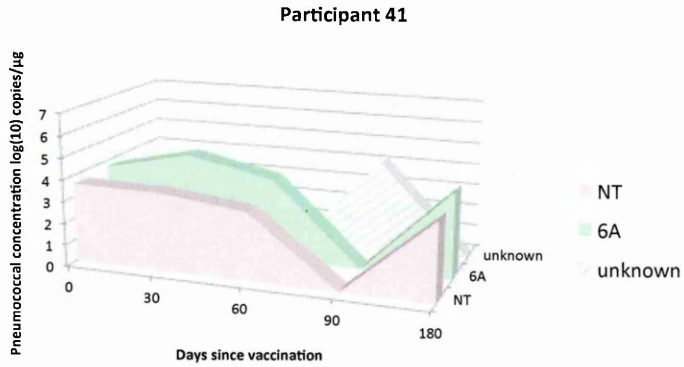
A


Participant 40
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15A
unknown
19A
20
11D
11D
20
19A
unknown
15A

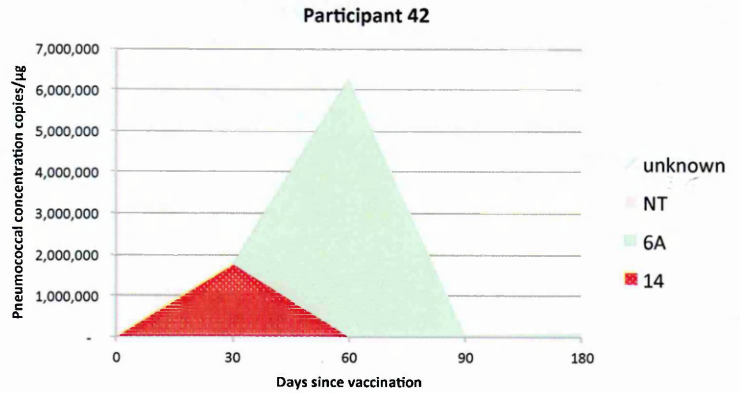
B


Participant 41
Pneumococcal concentration copies/µg
45,000
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
-
0
30
60
90
180
Days since vaccination
unknown
NT
6A

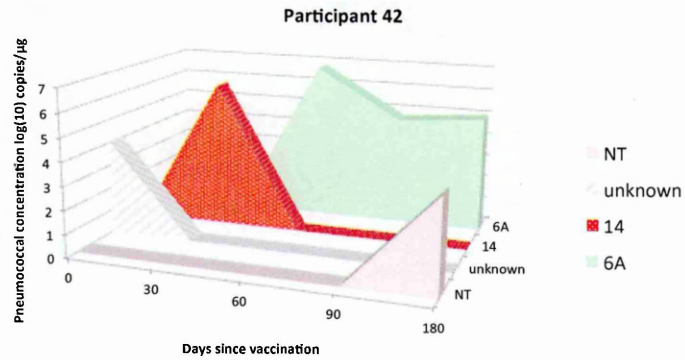
A


Participant 41
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
6A
NT
NT
6A
unknown

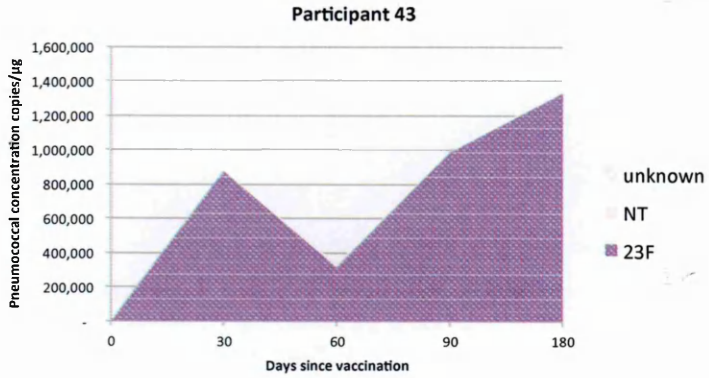
B


Participant 42
Pneumococcal concentration copies/µg
7,000,000
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
-
0
30
60
90
180
Days since vaccination
unknown
NT
6A
14

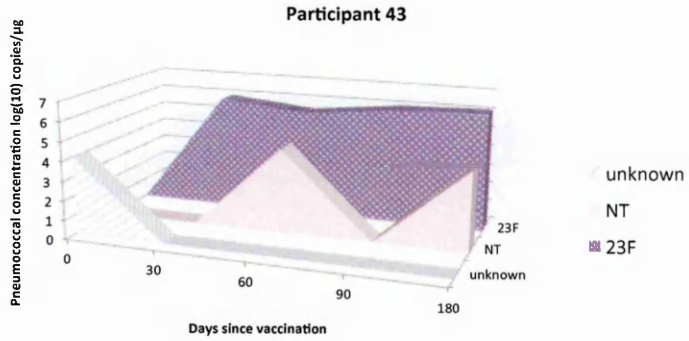
A


Participant 42
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
6A
14
unknown
NT
NT
unknown
14
6A

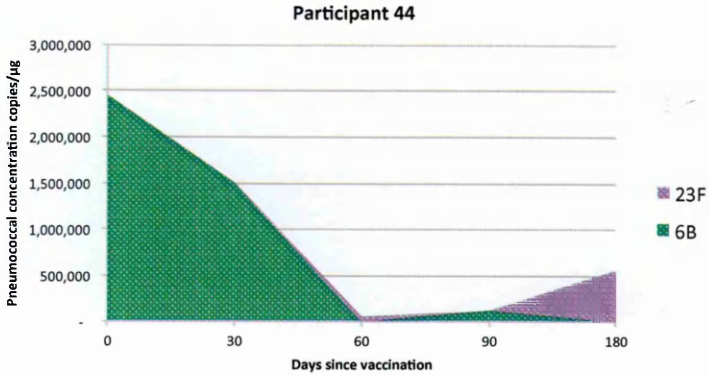
B


Participant 43
Pneumococcal concentration copies/µg
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
-
0
30
60
90
180
Days since vaccination
unknown
NT
23F

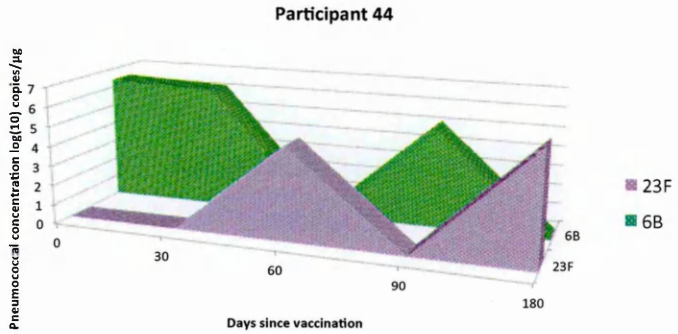
A


Participant 43
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
23F
NT
unknown
unknown
NT
23F

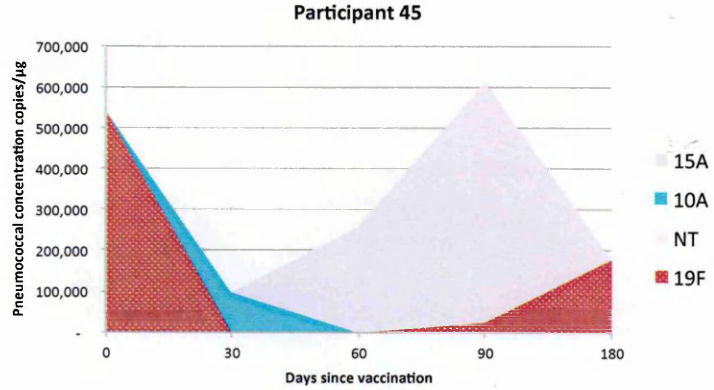
B


Participant 44
Pneumococcal concentration copies/µg
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
0
30
60
90
180
Days since vaccination
23F
6B

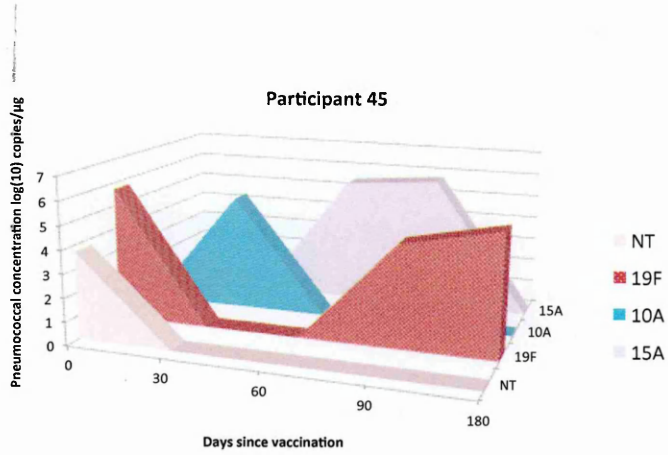
A


Participant 44
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
6B
23F
23F
6B

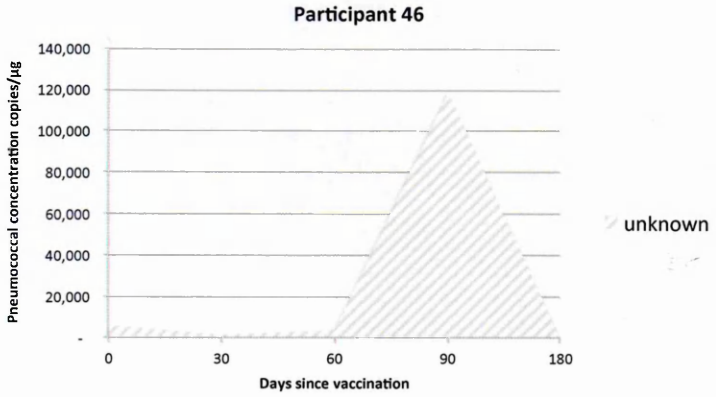
B


Participant 45
Pneumococcal concentration copies/µg
700,000
600,000
500,000
400,000
300,000
200,000
100,000
-
0
30
60
90
180
Days since vaccination
15A
10A
NT
19F

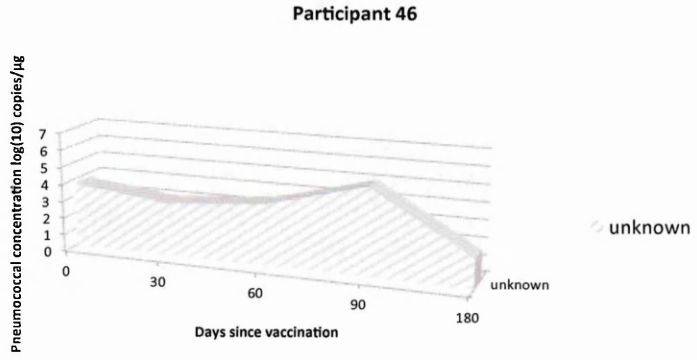
A


Participant 45
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
15A
10A
19F
NT
NT
19F
10A
15A

B


Participant 46
Pneumococcal concentration copies/µg
140,000
120,000
100,000
80,000
60,000
40,000
20,000
-
0
30
60
90
180
Days since vaccination
unknown

A


Participant 46
Pneumococcal concentration log(10) copies/µg
7
6
5
4
3
2
1
0
0
30
60
90
180
Days since vaccination
unknown
unknown

B

Appendix C: Nasopharyngeal pneumococcal concentration over six months post vaccination with 10-valent pneumococcal conjugate vaccine (PCV-10) by individual study participant in 46 children. Where swabs failed to grow on culture for microarray serotyping, thus serotype data is not available, the pneumococcal concentration of the unknown serotype/s is marked in grey stripes. PCV-10 vaccine-types (1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, 23F) are marked in stippled colours, non-vaccine-types are in plain colours. NT = not typable. Each study participant has two graphs; A and B. **A:** Pneumococcal concentration in *lytA* copies/µg of human DNA. Note that the y-axes are all to a different scale, due to the wide variation in pneumococcal concentration. Many of the minor serotypes cannot be seen due to the dominance of the major serotype, which can clearly be seen. **B:** Log_{10}-transformed pneumococcal concentration in log_{10} *lytA* copies/µg of human DNA. The same data as in the A graphs, log10-transformed, reveals the diversity of multiple serotype carriage.