Childhood Pneumococcal Pneumonia in Nepal



Michael James Carter

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

Pour mes co-pilotes: Stéphanie, Aurélien et Margot.

Acknowledgements and declarations

I would like to thank my three supervisors at the University of Oxford for their wisdom, support and guidance: Dr Dominic Kelly and Prof Andrew Pollard (Oxford Vaccine Group, Department of Paediatrics) and Prof Julian Knight (J Knight Group, Nuffield Department of Medicine). I would also like to thank Prof Shrijana Shrestha (Patan Academy of Health Sciences, Nepal; PAHS) for her guidance and the opportunity to work at PAHS. I would also like to thank members of Oxford Vaccine Group, the J Knight Group, the Department of Paediatrics at PAHS, and the PneumoNepal Project for their support.

This thesis contains data derived from three studies: The Childhood Invasive Bacterial Disease Surveillance Programme at Patan Hospital (March 2005 to December 2016), the Inpatient Pneumonia Study (March 2014 to December 2016) and the Pneumonia Aetiology in Nepal: Diagnosis with Immunology and Transcriptomics (PANDIT) Study (February 2015 to December 2016).

The Childhood Invasive Bacterial Disease Surveillance Programme

The Childhood Invasive Bacterial Disease Surveillance Programme was founded in 2005 as a collaboration between the Department of Paediatrics at PAHS, Oxford Vaccine Group, and the Department of Pathology at the University of Otago Christchurch, New Zealand. At PAHS, I would like to specifically acknowledge Dr Stephen Thorson, Prof Neelam Adhikari, Prof Shrijana Shrestha, Mr Krishna Prajapati, Dr Meeru Gurung and Dr Imran Ansari. In addition, I would like to acknowledge data support from Ruby and Bobby Basi, and the Pediatric Research Study and Microbiology Laboratory teams. These teams initially conceived the Programme and has managed the work on a daily basis over several years. At Oxford Vaccine Group, I would like to specifically acknowledge Dr Dominic Kelly and Prof Andrew Pollard who, with others, initially conceived the Programme and provide ongoing support. At the University of Otago Christchurch, Prof David Murdoch also initially conceived the Programme and provides ongoing support, particularly to the Microbiology Laboratory at PAHS.

Funding for the Programme was initially from the Pneumococcal vaccines Accelerated Development and Introduction Plan (PneumoADIP) through the South Asian Pneumococcal Alliance. The Programme receives ongoing support from the WHO Global Invasive Bacterial Vaccine-Preventable Disease Network, and Gavi, the Vaccine Alliance. My

specific role on the project was to collate, clean and analyse clinical and microbiological data from several databases arising from the studies. Analyses within this thesis are largely related to children with invasive pneumococcal disease.

Data from this programme relating to children enrolled between 2005 and 2006 have been previously published[4, 5], as have data from this programme relating to children enrolled between 2007 and 2009[6]. Data from 2010–2016 have not been previously published.

The Inpatient Pneumonia Study at Patan Hospital

The Inpatient Pneumonia Study commenced enrolment in March 2014 as a collaboration between PAHS, the University of the Oxford, the University of Otago Christchurch (members as above) and the International Vaccine Access Center (IVAC), Bloomberg School of Public Health, Johns Hopkins University, USA; and Agénce de Médecine Préventive (AMP), Paris, France. Additional members at PAHS included Dr Puja Amatya (radiological support), Bishnu Devi and Madhav Chandra Gautam (laboratory support). Additional members at the University of Oxford included Dr Rama Kandasamy (laboratory support), Dr Matthew Smedley and Dr Kate Park (radiological support), Sarah Kelly (project management) and Merryn Voysey and Dr Ly-Mee Yu (statistical support). Additional members at IVAC included Julie Yousskin (communications support) and Prof Kate O'Brien (study conception and design and epidemiological support). Additional members at AMP included Dr Jennifer Moïsi (study conception and design and epidemiological support) and Dr Bradford Gessner (study conception and design and epidemiological support).

The Inpatient Pneumonia Study was part of a larger programme of work, the Nepal Pneumococcal Impact Assessment (the PneumoNepal Project), funded by Gavi, the Vaccine Alliance. My specific role on the project was laboratory support (particularly quality control of Quellung serotyping), conduct and leadership of radiological analyses, and epidemiological analyses of data arising from the period March 2014 to December 2016. In particular, Merryn Voysey gave valuable advice on interrupted time series methods for this study.

The PANDIT Study at Patan Hospital

The PANDIT Study commenced enrolment in February 2015. The study was conceived and designed by me, Dr Kelly, Prof Pollard and Prof Knight. I enrolled and sampled the majority of patients enrolled into the PANDIT Study with support from the Clinical Research Fellow team, and Shristy Rajkarnikar and Pallavi Gurung (PANDIT Study research nurses). I processed the majority of blood samples for ALS in the Microbiology Laboratory at PAHS with the support of Shristy Rajkarnikar and Pallavi Gurung. I also thank Dr Claire Jones at Oxford Vaccine Group for training in this method, and for advice on the development of the Pneumovax ELISA. Laboratory analyses for IgG in ALS to

pneumococcal capsular polysaccharides was undertaken at the Dutch National Institute of Public Health and the Environment (RIVM), Bilthoven, led by Dr Guy Berbers with additional analyses at the WHO Pneumococcal Serology Reference Laboratories, UCL, London, led by Prof David Goldblatt. Laboratory analyses for IgG in ALS to pneumococcal proteins was undertaken at the Finnish National Institute for Health and Welfare (THL), Helsinki, led by Dr Merit Melin and Dr Nina Ekström. Dr Tim James at Oxford University Hospitals NHS Trust kindly allowed me to use space, reagents and equipment in the Biochemistry Laboratory to measure C-reactive protein in plasma samples from the Study.

On return to the University of Oxford, I extracted RNA from all samples in the study. RNA sequencing was at the Wellcome Trust Centre for Human Genetics, Oxford (supported by a Wellcome Trust grant 090532/Z/09/Z), led by Dr Jerome Nicod. I processed raw sequencing data with support from Dr Wanseon Lee and Dr Gabriele Migliorini (J Knight Group). My analyses of the mapped gene reads were supported by Dr Helen Lockstone (WTCHG), Dr Christoph Blohmke (OVG) and Katie Burnham (J Knight Group).

The PANDIT Study was funded by the Wellcome Trust via my Research Training Fellowship (grant 104439/Z/14/Z). Additional funding for RNA-sequencing of samples was through the Personalised Risk Assessment in Febrile Illness to Optimise Real-Life Management across the European Union (PERFORM) Study.

I specifically thank Shristy Rajkarnikar, Pallavi Gurung, Brian Wahl (IVAC), Dr Christoph Blohmke and Katie Burnham for exceptional camaraderie during these studies. I also thank our excellent friends and hosts in Nepal for their kindness and generosity towards the Carter *famille*.

I declare that unless otherwise stated, all work presented in this thesis is my own.

Submitted abstracts and publications

Submitted abstracts

Carter, M. J., Gurung, M., Prajapati, K. G., Murdoch, D. R., Adhikari, N., Ansari, I., Kelly, D. F., Thorson, S., Pollard, A. J. and Shrestha, S. (2016). *Invasive pneumococcal disease at Patan Hospital, Nepal: 2005–2016.* Poster presented at the *36th Annual Meeting of the European Society for Paediatric Infectious Diseases*, Madrid, Spain.

Carter, M. J., Rajkarnikar, S., Gurung, P., Gautam, M. C., Gurung, M., Ekström, N., Melin, M., Knight, J. C., Murdoch, D. R., Thorson, S., Shrestha, S., Pollard, A. J. and Kelly, D. F. (2017). *Antibodies from lymphocyte supernatant against pneumococcal capsular polysaccharides in children admitted to hospital with pneumonia in Nepal.* Poster presented at the *36th Annual Meeting of the European Society for Paediatric Infectious Diseases*, Madrid, Spain.

Carter, M. J., Gurung, M., Kandasamy, R., Gautam, M. C., Prajapati, K. G., Kelly, S., Wahl, B. P., Moïsi, J. C., Gessner, B. D., O'Brien, K. L., Murdoch, D. R., Shrestha, S., Kelly, D. F., Thorson, S. and Pollard, A. J. (2017). *Carriage of pneumococcal serotypes in community children and children with pneumonia in Nepal: preliminary results before and immediately following introduction of pneumococcal conjugate vaccination (PCV10)*. Poster presented at the *36th Annual Meeting of the European Society for Paediatric Infectious Diseases*, Madrid, Spain.

Carter, M. J., Gurung, M., Prajapati, K. G., Basi, R., Khadkha, B., Khadkha, B., Yadav, J., Pandit, R. Kelly, D. F., Pollard, A. J., Murdoch, D. R., Adhikari, N., Thorson, S., Ansari, I. and Shrestha, S. (2016). *Bacterial aetiology of meningitis in children from Kathmandu, Nepal, 2005–2014.* Poster presented at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Amatya, P., **Carter, M. J.**, Smedley, M., Gurung, M., Kelly S., Khadkha B., Khadkha B., Murdoch D. R., Ansari I., Thorson S., Shrestha S., Kelly D. F., Pollard A. J. and Park K. (2016). *Incidence of WHO-defined radiographic primary endpoint pneumonia in children with clinically diagnosed pneumonia at Patan Hospital, Nepal.* Oral presentation at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Ansari, I., **Carter, M. J.**, Gurung, M., Basi, R., Prajapati, K. G., Adhikari, N., Shrestha, S., Murdoch, D. R., Kelly, D. F., Pollard, A. J. and Thorson, S. (2016). *Pneumonia in children admitted to Patan Hospital, Nepal, 2005–2014: differing clinical criteria and the surveyed prevalence of pneumonia.* Poster presented at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Kandasamy, R., **Carter, M. J.**, Shrestha, S., Ansari, I., Gautam, M. C., Ndimah, S., Thorson, S., O'Brien, K. L., Kelly, S., Murdoch, D. R., Kelly, D. F. and Pollard A. J. (2016). *Comparison of childhood pneumococcal carriage and serotype distribution between inpatients with pneumonia and healthy controls in Nepal prior to pneumococcal conjugate vaccine introduction.* Poster presented at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Shrestha, S., **Carter, M. J.**, Gurung, M., Kandasamy, R., Adhikari, N., Prajapati, K. G., Kelly, S., Kelly, D. F., Pollard, A. J., Ansari, I., Thorson, S. and Murdoch, D. R. (2016). *Bacteraemia in children with suspected invasive bacterial disease at Patan Hospital, Nepal, 2005–2015: age distribution and time series analysis.* Poster presented at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Gurung, M., Kandasamy, R., **Carter, M. J.**, Khadkha, B., Maharjan, A., Khadka, B., Prajapati, K. G., Shrestha, S., Ansari, I., Adhikari, N., Thorson, S., Murdoch, D. R., Kelly, D. F. and Pollard A. J. (2016). *Childhood pneumococcal disease epidemiology at Patan Hospital, Nepal 2005–2015: invasive disease and serotype distribution.* Poster presented at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Wahl, B., Gurung, M., Thorson, S., Sharma, A., Shrestha, S., **Carter, M. J.**, Kelly, S., O'Brien, K. L., Gessner, B. D., Pollard, A. J., Ansari, I. and Moïsi, J. C. (2016). *Measuring the impact of 10-valent pneumococcal conjugate vaccine on pneumonia, meningitis and sepsis hospitalizations in Nepal: proposed methods.* Poster presented at the *International Symposium on Pneumonia and Pneumococcal Disease-10*, Glasgow, UK.

Publications arising directly from this work

Carter, M. J., Mitchell, R. M., Meyer Sauteur, P. M., Kelly, D. F. and Trück, J. (2017). The antibody-secreting cell response to infection: kinetics and clinical applications. *Front Immunol* 8, pp1664–3224.

Other publications

Darton, T. C., Baker, S., Randall, R., Dongol, S., Karkey, A., Voysey, M., **Carter, M. J.**, Jones, C., Trappl, K., Pablo, J., Hung, C., Teng, A., Shandling, A., Le, T., Walker, C., Andrews, J., Arjyal, A., Basnyat, B., Pollard, A. J. and Blohmke C. J. (Submitted).

Identification of new serodiagnostic signatures of acute typhoid fever using a *Salmonella* proteome array.

Darton, T. C., Jones, C., Dongol, S., Voysey, M., Blohmke, C. J., Shrestha, R., Karkey, A., Shakya, M., Arjyal, A., Waddington, C., Gibani, M., **Carter, M. J.** Basnyat, B., Baker, S. and Pollard, A. J. (In review). An assessment of the Antibodies in Lymphocyte Supernatant (ALS) assay for the diagnosis of enteric fever in two human challenge models in an endemic area of Nepal.

Herberg, J. A., Kaforou, M., Wright, V. J., Shailes, H., Eleftherohorinou, H., Hoggart, C. J., Cebey-Lopez, M., **Carter, M. J.**, Janes, V. A., Gormley, S., Shimizu, C., Tremoulet, A. H., Barendregt, A. M., Salas, A., Kanegaye, J., Pollard, A. J., Faust, S. N., Patel, S., Kuijpers, T., Martinon-Torres, F., Burns, J. C., Coin, L. J., Levin, M. and the Iris Consortium. (2016). Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. *J Am Med Assoc* 316(8), pp.835–45.

Carter, M. J., Blohmke, C. J. and Pollard, A. J. (2016). *Immunological correlates of vaccine-mediated protection*. In: *The Vaccine Book, Second Edition*. Editors: Bloom B. R. and Lambert P. H. Elsevier, London.

Emary, K. R., **Carter, M. J.**, Pol, S., Sona, S., Kumar, V., Day, N. P., Parry, C. M. and Moore, C. E. (2015). Urinary antibiotic activity in paediatric patients attending an outpatient department in north-western Cambodia. *Trop Med Int Health* 20(1), pp24–8.

Carter, M. J., Emary, K. R., Moore, C. E., Parry, C. M., Sona, S., Putchhat, H., Reaksmey, S., Chanpheaktra, N., Stoesser, N., Dobson, A. D., Day, N. P., Kumar, V., Blacksell, S. D. (2015). Rapid diagnostic tests for dengue virus infection in febrile Cambodian children: diagnostic accuracy and incorporation into diagnostic algorithms. *PLoS Negl Trop Dis* 9(2): e0003424.

Abstract

Pneumonia is the greatest cause of childhood mortality outside the neonatal period[1], yet the pathogen-specific aetiology of childhood pneumonia remains poorly defined[2]. Vaccine probe studies estimate that approximately one third of children <5 years of age with radiographic endpoint consolidation have pneumococcal pneumonia in settings prior to the introduction of pneumococcal conjugate vaccination (PCV)[3], such as much of South Asia. 10-valent PCV was introduced to the Nepali infant immunisation schedule in August 2015.

I investigated childhood pneumococcal pneumonia in Nepal. I aimed to describe the prevalence of pneumococcal infection in children with suspected invasive bacterial disease admitted to Patan Hospital, Kathmandu, Nepal; to assess the impact of 10-valent PCV on pneumococcal pneumonia; and to assess two potential diagnostic tests for pneumococcal pneumonia based on the childhood response to pneumococci – assay of antibodies from lymphocyte supernatant (ALS) and analysis of differential gene expression (transcriptomics) – in children with pneumonia at Patan Hospital.

Pneumococci were the second-most most prevalent pathogen isolated from the blood of children between 2005 and 2016. Interrupted time series analyses of data from children admitted with pneumonia from March 2014–December 2016, showed a small increase ($\sim$4%) in the odds of admission with pneumonia in comparison to non-pneumonia admissions temporally associated with 10-valent PCV introduction. However, it was not possible to adjust these time series analyses for extreme events including earthquakes (April 2015) and clean fuel shortages/increased air pollution (winter 2015/2016). In contrast, the indirect cohort method (a case-control approach in vaccinated vs unvaccinated children) showed vaccine effectiveness of 84% on the odds of nasopharyngeal carriage of vaccine-type pneumococci, but no effectiveness on pneumonia in these early data. Assay of IgG ALS pneumococcal capsular polysaccharides was complicated by what appear to be non-specific binding to capsular polysaccharides (of both *S. pneumoniae*, particularly serotype 3, and *H. influenzae)*. Assay of IgG ALS to the best-performing of five pneumococcal proteins assessed had a sensitivity of 88% and specificity of 71% for the discrimination of pneumococcal pneumonia from other bacterial pneumonia. A transcriptomic signature discriminated between pneumococcal pneumonia and other bacterial pneumonia with a sensitivity of 91% and specificity of 100%

This thesis therefore contributes to knowledge of the clinical epidemiology of pneumococcal disease in South Asia. These data may also contribute to public health policy-making in the region. In addition, the development of two diagnostic tests for the aetiology of childhood pneumonia may be useful for future studies of childhood pneumonia aetiology.

Contents

List of Figures **xv**

List of Tables **xx**

List of Abbreviations **xxiii**

1 Introduction: The diagnosis of pneumonia and pneumococcal disease in children **1**

1.1 Pneumonia in children 1

1.1.1 Clinical features of pneumonia, meningitis and sepsis in children . 4

1.2 Pneumococcal disease in children 9

1.2.1 *Streptococcus pneumoniae* as bacterium, coloniser and pathogen 9

1.2.2 Diagnostic methods for childhood pneumococcal pneumonia . . . 13

1.2.3 Vaccine probes to estimate childhood pneumococcal pneumonia . 21

1.3 Novel diagnostic tests for the aetiology of pneumonia 22

1.3.1 The ASC response to infection 23

1.3.2 Measuring ASCs in infection 25

1.4 The transcriptomic response to infection 38

1.4.1 Transcriptomic techniques 39

1.4.2 Transcriptomics for the diagnosis of infectious diseases 40

1.5 Aim and objectives 42

2 General methods **45**

2.1 Study site and population 45

2.1.1 Introduction of PCV to the Nepali infant immunisation schedule . . 48

2.1.2 Overview of case definitions used within these studies 48

2.2 Childhood invasive bacterial disease surveillance 48

2.2.1 Definitions, inclusion criteria and clinical methods 48

2.2.2 Microbiology: cultures from normally sterile sites 50

2.3 Childhood pneumonia studies 50

2.3.1 Definitions, inclusion criteria and clinical methods 51

2.3.2 Microbiology: cultures of nasopharyngeal swabs and serotyping . 53

2.3.3 Radiology: interpretation of chest radiographs 54
2.3.4 The Community Carriage Study 55
2.3.5 Data management and analysis 56
2.4 Methods specific to the PANDIT study 56
2.4.1 Methods to generate ALS 57
2.4.2 Antibodies to pneumococcal polysaccharides by fluorescent multiplexed bead-based immuno-assay 58
2.4.3 Antibodies to pneumococcal proteins by fluorescent multiplexed bead-based immuno-assay 58
2.5 Transcriptomic investigation of childhood pneumonia: the PANDIT study . 58
2.5.1 Sampling, extraction and quality control of RNA 58
2.5.2 Bioinformatic methods for RNA-sequencing data 59

3 Childhood invasive pneumococcal disease and pneumonia at Patan Hospital 63
3.1 Introduction 63
3.2 Methods 66
3.2.1 Overview of studies 66
3.2.2 Invasive pneumococcal disease, April 2005–December 2016 . . . 67
3.2.3 Childhood pneumonia at Patan Hospital, March 2014–December 2016 68
3.3 Results 69
3.3.1 Invasive pneumococcal disease, April 2005–December 2016 . . . 69
3.3.2 Childhood pneumonia at Patan Hospital, March 2014–December 2016: 75
3.4 Discussion 93
3.4.1 Invasive pneumococcal disease 93
3.4.2 The Inpatient Pneumonia Study 97
3.4.3 Proxy markers of pneumococcal pneumonia 98
3.4.4 Strengths, limitations, future work and summary 101

4 Impact of pneumococcal conjugate vaccination in Nepal: preliminary assessment 103
4.1 Introduction 103
4.2 Methods 105
4.2.1 General methods 105
4.2.2 Interrupted time series analyses 106
4.2.3 Indirect cohort method 109
4.3 Results 110
4.3.1 Descriptive analyses of enrolled patients over time 110

4.3.2 Results of interrupted time series analyses . . . 113
4.3.3 Results of indirect cohort methods . . . 117
4.4 Discussion . . . 120
4.4.1 Interrupted time series analyses . . . 120
4.4.2 Indirect cohort method . . . 128
4.4.3 Summary and implications for future work . . . 129

5 ALS for the detection of pneumococcal responses from antibody-secreting cells 131
5.1 Introduction . . . 131
5.2 Methods . . . 133
5.2.1 Enrolment and sampling . . . 134
5.2.2 Development of comparator standards for pneumonia aetiology . . . 135
5.2.3 ALS methods . . . 137
5.2.4 FMIA to 13 pneumococcal capsular polysaccharides . . . 137
5.2.5 FMIA to 5 pneumococcal proteins . . . 139
5.2.6 Assays to control antigens and further investigations . . . 140
5.3 Results . . . 141
5.3.1 Enrolment and clinical classifications . . . 141
5.3.2 FMIA for the quantitation pneumococcal capsular polysaccharide specific IgG ALS . . . 141
5.3.3 Results of FMIA to detect ALS to pneumococcal proteins . . . 151
5.4 Discussion . . . 158
5.4.1 Pneumonia aetiology comparator standards . . . 160
5.4.2 IgG ALS has low specificity and sensitivity for pneumococcal capsular polysaccharides in children with pneumonia . . . 161
5.4.3 IgG ALS may discriminate between pneumonia aetiology comparator standards . . . 162
5.4.4 Assay of ALS as a novel diagnostic test for pneumonia aetiology? 163
5.4.5 Limitations and summary . . . 164

6 Transcriptomic investigation of childhood pneumonia 166
6.1 Introduction . . . 166
6.2 Methods . . . 168
6.2.1 Enrolment and sampling . . . 169
6.2.2 Development of comparator standards for pneumonia aetiology . . . 170
6.2.3 Peripheral blood RNA extraction and quantification . . . 170
6.2.4 Bioinformatic approaches . . . 171
6.3 Results . . . 174
6.3.1 Enrolment, sampling and comparator standards . . . 174

6.3.2 Principal component analyses 177
6.3.3 Comparison 1: all-cause pneumonia vs healthy controls 181
6.3.4 Comparison 2: acute pneumonia vs convalescence 181
6.3.5 Comparison 3: bacterial pneumonia vs viral pneumonia 190
6.3.6 Comparison 4: pneumococcal pneumonia vs other bacterial pneumonia 196
6.4 Discussion 200
6.4.1 The transcriptome discriminates between healthy controls and disease 202
6.4.2 Immunological pathways are upregulated in children with pneumonia 207
6.4.3 A distinct transcriptomic signature that discriminates between bacterial and viral pneumonia 208
6.4.4 A distinct transcriptomic signature that discriminates between pneumococcal and other bacterial pneumonia 210
6.4.5 Limitations and summary 211

7 Summary: childhood pneumococcal pneumonia at Patan Hospital, Nepal 213
7.1 Pneumococci are the most common invasive pathogens in childhood pneumonia 215
7.2 Early impact of 10-valent PCV in children with pneumonia 216
7.3 IgG ALS to pneumococcal proteins for aetiological diagnosis 217
7.4 Transcriptomics for aetiological diagnosis 219
7.5 Implications for future studies of pneumonia aetiology 220

Appendices

A Appendix 223
A.1 General methods 223
A.2 Transcriptomic investigation of childhood pneumonia 223
A.2.1 Bioinformatic methods: scripts 223
A.2.2 Mapping of reads 223
A.2.3 Filtering of reads 225
A.2.4 Generation of gene counts and edgeR pipeline 225

List of Figures

1.1 Historical pneumococcal conjugate vaccine (PCV) introduction by income group. 3
1.2 Annual under-5 all-cause and pneumonia mortality rates with global targets and pivotal health interventions, by time. 4
1.3 Differentiation of B cells in response to antigen. 26

2.1 Overview of time course for enrolment, and inclusion criteria, for studies described in this thesis. 49
2.2 Examples of a "normal" radiograph, a radiograph with "other infiltrate only", and a radiograph with "endpoint consolidation" from children enrolled in the Inpatient Pneumonia Study at Patan Hospital. 55
2.3 Schematic overview of the bioinformatic processing of RNA samples from children enrolled in the study of the transcriptomic response to pneumonia. 61

3.1 Enrolment of children 0–14 years of age with community-acquired disease to the childhood IBD surveillance programme at Patan Hospital by year. . 71
3.2 Childhood IBD at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures by organism/group by year. 74
3.3 Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures for *Streptococcus pneumoniae* by age group of child year of admission. 76
3.4 Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures for *Streptococcus pneumoniae* by age of child and presence of serotype in 10-valent and 13-valent PCV. 77
3.5 Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures for *Streptococcus pneumoniae* by age of child and whether the isolate was of serotype 1. . 78
3.6 Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: relationship between pneumococcal serotype and presence in 10-valent and 13-valent PCV by year and by age group. 79

3.7 Nasopharyngeal carriage of serotypes of pneumococci in enrolled children <2 years of age (n = 611), 2 to 5 years of age (n = 215) and 5 to 14 years of age (n = 112) admitted to Patan Hospital with a clinician diagnosis of pneumonia, March 2014–December 2016. 88
3.8 Nasopharyngeal carriage of serotypes of pneumococci in children 6 months to 2 years of age (n = 2895) and 2 to 5 years of age (n = 307) in the community, according to age group, March 2014–December 2016. . . 89
3.9 Odds ratios for NP carriage of pneumococcal serotypes contained within 13-valent PCV and non-typeable (NT) pneumococci in pneumonia and community control children (all ages), adjusted for age and sex, prior to introduction of PCV10 into the Kathmandu Valley. 91
3.10 Classification tree for the prediction of invasive pneumococcal disease in children with pneumonia at Patan Hospital using the variables: absence of wheeze on examination, chest radiograph, blood neutrophil count x 10^3/mm^3, chest radiograph results, and NP carriage of pneumococcal serotypes. 94
3.11 Classification tree for the prediction of invasive pneumococcal disease in children with pneumonia at Patan Hospital using the variables: absence of wheeze on examination, chest radiograph, CRP (mg/l), blood neutrophil count x 10^3/mm^3, chest radiograph results, and NP carriage of pneumococcal serotypes. 94
3.12 Classification tree for the prediction of invasive pneumococcal disease in children with pneumonia at Patan Hospital using the variables: age of child, absence of wheeze on examination, chest radiograph, CRP (mg/l), blood neutrophil count x 10^3/mm^3, chest radiograph results, and NP carriage of pneumococcal serotypes. 95

4.1 Enrolment of children 2 months to 14 years of age with clinician diagnosed pneumonia by age group, by all ages, total non-pneumonia admissions, and all childhood admissions to Patan Hospital per month, March 2014–December 2016. 111
4.2 The proportion of children vaccinated with one or more, two or more, or three doses of 10-valent PCV, by age group. PCV10, 10-valent PCV. . . . 112
4.3 Enrolment of children 2 months to14 years of age with clinician-diagnosed pneumonia, according to clinical feature, by month of enrolment, March 2014–December 2016. 113
4.4 Clinical features of enrolled children under 5 years of age admitted to Patan Hospital with clinician diagnosed pneumonia by age group, March 2014–December 2016. 114

4.5 Characteristics of children 0 to 14 years of age with clinician-diagnosed pneumonia, according to radiographic classification, by month of enrolment, March 2014–December 2016. 115
4.6 Enrolment of children 0 to 14 years of age with clinician-diagnosed pneumonia, according to radiographic feature, by month of enrolment and by age group, March 2014–December 2016. 116
4.7 Enrolment of children 0 to 14 years of age with clinician-diagnosed pneumonia, according to prior treatment with antibiotics, by month of enrolment, March 2014–December 2016. 117
4.8 Enrolment of children 0 to 14 years of age with clinician diagnosed pneumonia, according to carriage of pneumococcal serotypes, by month of enrolment, March 2014–December 2016. 118
4.9 Enrolment of children 0 to 14 years of age with clinician diagnosed pneumonia, according to NP carriage of pneumococcal serotypes, by month of enrolment and age group, March 2014–December 2016. 119
4.10 Analysis of all pneumonia admissions over time in children <2 years of age in the Inpatient Pneumonia Study. 120
4.11 Analysis of clinical pneumonia phenotypes over time in children <2 years of age in the Inpatient Pneumonia Study. 121
4.12 Analysis of pneumonia with endpoint consolidation over time in children <2 years of age in the Inpatient Pneumonia Study. 122
4.13 Analysis of NP carriage of vaccine-type pneumococci / non-vaccine-type pneumococci by quarter of enrolment in children <2 years of age in the Inpatient Pneumonia Study. 122

5.1 Enrolment and sample availability for the study of ALS as a diagnostic test for the aetiology of pneumonia (within the PANDIT study). 142
5.2 Quantification of IgG ALS to pneumococcal capsular polysaccharides in 100 children with pneumonia in Nepal by order of enrolment into the study. 145
5.3 IgG ALS to pneumococcal capsular polysaccharides in 100 children with pneumonia in Nepal, ordered by levels of IgG to PS3. 146
5.4 IgG to pneumococcal capsular polysaccharides in ALS from 10 children with pneumonia in Nepal. 148
5.5 ALS to pneumococcal capsular polysaccharides (within 13-valent PCV) by pneumonia aetiology comparator standard 150
5.6 IgG ALS to serotype 1 capsular polysaccharide by pneumonia aetiology comparator standard 151
5.7 ROC curves to determine optimal thresholds for IgG ALS concentrations to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia for five pneumococcal proteins: CbpA, PcsB, PhtD, Ply and Stkp-C. 156

5.8 IgG ALS to pneumococcal proteins CbpA, PcsB, PhtD, Ply and Stkp-C by pneumonia aetiology comparator standard. 157
5.9 ROC curves to determine optimal thresholds for IgG ALS concentrations to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia in children ≥2 years of age for five pneumococcal proteins: CbpA, PcsB, PhtD, Ply and Stkp-C. 159
5.10 IgG ALS to pneumococcal proteins CbpA, PcsB, PhtD, Ply and Stkp-C by pneumonia aetiology comparator standard in children ≥2 years of age. . . 160

6.1 Enrolment and sample availability for the transcriptomic investigation of childhood pneumonia 175
6.2 Comparisons of differential gene expression to develop transcriptomic signatures for childhood pneumonia across comparator standards of pneumonia aetiology and healthy controls. 176
6.3 Variance in gene expression across cases (acute and convalescent samples) and healthy controls shown by principal component analysis. 179
6.4 Principal component 1 and 2 of all differential gene expression between children with acute pneumonia (all comparator standards) and healthy controls. 180
6.5 Volcano plot of differential gene expression between children with acute pneumonia (any comparator standard) and healthy controls. 182
6.6 Heat map of differential gene expression between children with acute pneumonia (any comparator standard) and healthy control children. . . . 183
6.7 Enrichment of gene networks between children with acute pneumonia (any comparator standard) and healthy control children. 184
6.8 Principal component 1 and 2 of all differential gene expression between a child with acute pneumonia, and the same child in convalescence (6–8 weeks following the first sample). 185
6.9 Volcano plot of differential gene expression between children with acute pneumonia (any comparator standard) and a paired convalescent sample (6–8 weeks following enrolment). 186
6.10 Heat map of differential gene expression between children with acute pneumonia (any comparator standard) and a paired convalescent sample (6–8 weeks following enrolment). 187
6.11 Enrichment of gene networks between children with acute pneumonia (any comparator standard) and a paired convalescent sample (6–8 weeks following enrolment). 188
6.12 Volcano plot of differential gene expression between children with pneumococcal pneumonia and a paired convalescent sample (6–8 weeks following enrolment). 190

6.13 Heat map of differential gene expression between children with pneumococcal pneumonia and a paired convalescent sample (6–8 weeks following enrolment). 191
6.14 Enrichment of gene networks between children with pneumococcal pneumonia and a paired convalescent sample (6–8 weeks following enrolment). 192
6.15 Volcano plot of differential gene expression between children with bacterial pneumonia and viral pneumonia. 193
6.16 Heat map of differential gene expression between children with bacterial pneumonia and viral pneumonia. 194
6.17 Enrichment of gene networks between children with bacterial pneumonia and *probable viral* pneumonia. 195
6.18 Application of a transcriptomic signature to discriminate between bacterial and viral pneumonia, the Bacterial Gene Score. 196
6.19 ROC curve of a transcriptomic signature to discriminate between bacterial and viral pneumonia, the Bacterial Gene Score. 197
6.20 Boxplots showing the Bacterial Gene Score applied to children with pneumonia of all comparator standards and healthy controls. 199
6.21 Volcano plot of differential gene expression between children with pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score. 200
6.22 Heat map of differential gene expression between children with pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score. 201
6.23 Boxplots describing a transcriptomic signature to discriminate between pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score. 202
6.24 ROC curve of a transcriptomic signature to discriminate between pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score. 203
6.25 Boxplots showing the Pneumococcal Gene Score applied to children with pneumonia of all comparator standards and healthy controls. 206

A.1 Methods for the generation of ALS in the pneumonia studies at Patan Hospital. 224

List of Tables

1.1 Clinical features of pneumonia, meningitis and sepsis as determined by the WHO *IMCI* guidelines[18] and *Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illness*[36]. 8
1.2 Summary of diagnostic tests used in the PERCH study for the aetiological diagnosis of childhood pneumonia. 15
1.3 Studies investigating the use of ELISpot and ALS for the diagnosis of infection. 29

2.1 Standarised interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies[217] 54
2.2 Media used for the separation of PBMCs and the generation of ALS. . . . 62

3.1 Demographic characteristics of children enrolled into the childhood IBD surveillance programme, and of children with invasive pneumococcal disease. 72
3.2 Number (and percentage) of children admitted to Patan Hospital with suspected IBD, with bacteria isolated from blood, CSF or pleural fluid, by year. 73
3.3 Characteristics of patients enrolled in the study by age group. 80
3.4 Characteristics of patients in the study by presence or absence of wheeze on admitting examination. 82
3.5 Numbers of admission radiographs by radiographic classification and details on the concordance between readers (A) and each reader and the arbiter (B and C) on the diagnosis of endpoint consolidation on admission chest radiographs from children enrolled in the study. 83
3.6 Characteristics of patients enrolled in the study by radiographic classification 85
3.7 Characteristics of patients enrolled in the study by carriage of vaccine-type (10-valent PCV), non-vaccine-type or no NP carriage of pneumococci. . . 87
3.8 Association of NP carriage of pneumococcal serotypes within 10-valent and 13-valent PCV with absence of wheeze, and blood neutrophil count in children enrolled to the Inpatient Pneumonia Study at Patan Hospital. . . 90

3.9 Odds ratios for NP carriage of pneumococcal serotypes contained within 13-valent PCV and non-typeable (NT) pneumococci in pneumonia and community control children, adjusted for age and sex, *prior* to the introduction of 10-valent PCV into routine infant immunisation schedule of the Kathmandu Valley. 91
3.10 Odds ratios (OR) for association between clinical features, laboratory results, and NP carriage of pneumococcal serotypes, unadjusted and adjusted for age of child for invasive pneumococcal disease (IPD). 92

4.1 Outcomes variables for interrupted time series analyses 108
4.2 Interrupted time series analyses of changes in the number of enrolled children under 2 years of age with clinical features (fast breathing, chest indrawing, danger signs, absence of wheeze or radiographic endpoint consolidation / non-pneumonia admissions in this age group), and NP carriage of vaccine-type pneumococci / carriage of non-vaccine-type pneumococci associated with increasing proportion children vaccinated with 10-valent PCV. 115
4.3 Characteristics of patients enrolled in the study, and eligible for two or more doses of 10-valent PCV through routine immunisation, by vaccination status. The analyses of vaccine effectiveness using the indirect cohort method was done using data from this subset of the case series. 123
4.4 Indirect cohort method: estimates of vaccine effectiveness against various outcome measures in children eligible for two or more doses of 10-valent PCV. 124

5.1 Characteristics of children on whom IgG ALS to pneumococcal capsular polysaccharides was quantified by pneumonia aetiology comparator standard. 144
5.2 IgG concentrations (μg/ml) as measured by FMIA to purified capsular polysaccharides 1 and 3 following addition of ALS with 50 μg/ml of purified capsular polysaccharides (PS) of serotypes 1 and 3. 147
5.3 Concentrations of *Haemophilus influenzae* (Hib) type B specific IgG ALS from patients with pneumonia. 149
5.4 IgG as measured by FMIA to pertussis antigens, diphtheria toxin and tetanus toxoid in ALS. These children are from a population with high vaccine coverage for diphtheria-tetanus-pertusiss. 149
5.5 Characteristics of children on whom ALS to pneumococcal proteins was measured by pneumonia aetiology comparator standard. 153
5.6 IgG ALS to the pneumococcal proteins CbpA, PcsB, PhtD, Ply and Stkp-C by pneumonia aetiology comparator standard 154

5.7 Optimal thresholds and diagnostic accuracy (95% CIs) measures for IgG ALS concentrations to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia for five pneumococcal proteins: CbpA, PcsB, PhtD, Ply and Stkp-C. 155

6.1 Characteristics of children on whom transcriptomic investigations were undertaken, by pneumonia aetiology comparator standard. 178
6.2 Characteristics of children with *definite pneumococcal* (DP), PP, *probable pneumococcal* (PP) and *definite other bacterial* (DOB) pneumonia on whom transcriptomic investigations were undertaken, by pneumonia aetiology comparator standard. 189
6.3 Diagnostic accuracy of the Bacterial Gene Score, using a threshold derived from ROC curve analysis of bacterial vs viral pneumonia. 198
6.4 Diagnostic accuracy of the Pneumococcal Gene Score, using a threshold derived from ROC curve analysis of pneumococcal vs *definite other bacterial* gene score. 204
6.5 Genes differentially expressed in both a comparison of acute pneumococcal pneumonia and convalescence and in a comparison of pneumococcal pneumonia and *definite other bacterial* pneumonia. 205

List of Abbreviations

ALS . . . . . Antibodies from lymphocyte supernatant

AMP . . . . . Agénce de Médecine Préventive, Paris, France

ASC . . . . . Antibody-secreting cell

BCG . . . . . Bacille Calmette-Guérin

BCR . . . . . B cell receptor

CbpA . . . . . Choline binding protein A

CD . . . . . Cluster of differentiation (on a B cell membrane)

CD40L . . . . . Ligand for CD40

CFA/II . . . . . *Escherichia coli* colonising factor

CLSI . . . . . Clinical and Laboratory Standards Institute (Wayne, PA, USA)

CPS . . . . . Capsular polysaccharide (gene in lower case and italicised)

CRP . . . . . C-reactive protein

CS1, CS3, CS6 . Coli surface antigens (from *E. coli*)

CSF . . . . . Cerebrospinal fluid

CtxB . . . . . Cholera toxin B subunit

CWPS . . . . . Cell wall polysaccharide (pneumococcal; techoic acid)

DOB . . . . . Definite other bacterial (pneumonia)

DP . . . . . Definite pneumococcal (pneumonia)

ELISA . . . . . Enzyme-linked immunoassay

ELISpot . . . . . Enzyme-linked immunspot

EPI . . . . . Extended Programme on Immunisation

ETEC . . . . . Enterotoxigenic *E. coli*

EU . . . . . Union européene/European Union

EV-71 . . . . . Enterovirus-71

FDR . . . . . False discovery rate (derived from the Benjamini-Hochberg procedure)

GMC . . . . . Geometric mean concentration

HBOV . . . . . Human bocavirus

HBV . . . . . Hepatitis B virus

Hib . . . . . *Haemophilus influenzae* type b

HMPV . . . . . Human metapneumovirus

HPV . . . . . Human papillomavirus

IMCI . . . . . Integrated Management of of Childhood Illness (WHO guidelines)

IPD . . . . . Invasive pneumococcal disease

IVAC . . . . . International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

LLPCs . . . . . Long-lived plasma cells

LPS . . . . . Lipopolysaccharide

LT . . . . . Heat-labile *E. coli* toxin

LytA . . . . . Autolysin (gene in lower case and italicised)

MSHA . . . . . Mannose-sensitive haemogluttinin

MSP1 . . . . . Merozoite surface protein 1

NIBSC . . . . . National Institute for Biological Standards and Control (Herts, UK)

NP . . . . . Nasopharyngeal (swab or carriage, i.e. of pneumococci)

OmpA . . . . . Outer membrane protein A

OVG . . . . . Oxford Vaccine Group

PERCH . . . . . Pneumonia Etiology Research for Child Health

PB . . . . . Probable bacterial (pneumonia)

PBMC . . . . . Peripheral blood mononuclear cells

PBS . . . . . Phosphate buffer saline

PCR . . . . . Polymerase chain reaction

PCV . . . . . Pneumococcal conjugate vaccine (currently licensed vaccines are the 10-valent PCV (*Synflorix*, GSK), and 13-valent PCV (*Prevnar 13*, Pfizer)

PcpA . . . . . Pneumococcal choline binding protein A

PcsB-N . . . . . Protein for cell wall separation of group B streptococci

PhtD . . . . . Pneumococcal histidine triad protein D

Ply . . . . . Pneumolysin protein (gene in lower case and italicised)

PP . . . . . Probable pneumococcal (pneumonia)

PPD . . . . . Purified protein derivative

PS . . . . . Polysaccharide (PS)

PsaA . . . . . Pneumococcal surface adhesion molecule (gene in lower case and italicised)

PV . . . . . Probable viral (pneumonia)

PV/EV . . . . . Parechovirus/enterovirus

qPCR . . . . . Quantitative polymerase chain reaction

R0 . . . . . Cell culture medium consisting of RPMI 1640 with the addition of penicillin, streptomycin, and L-glutamine. L-glutamine was replenished every 14 days.

R10 . . . . . Cell culture medium consisting of R0 with the addition or 10% fetal calf serum.

RIVM . . . . . Rijksinstitut voor Volksgezondheid en Milieu (National Institute for Public Health and the Environment, Bilthoven, Netherlands)

ROC . . . . . Receiver operating characteristics (curves)

RPM . . . . . Revolutions per minute (of a centrifuge)

RPMI . . . . . Roswell Park Memorial Institute medium (RPMI medium 1640 is used for mammalian cell culture)

RSV . . . . . Respiratory syncytial virus

STGG . . . . . Skim-milk-tryptone-glucose-glycerin (medium for culture of pneumococci from nasopharyngeal swabs)

Stkp-C . . . . . Serine/threonine protein kinase-C

TB . . . . . Tuberculosis

TCR . . . . . T cell receptor

Th cell . . . . . T helper cell

THL . . . . . Terveyden ja hyvinvoinnin laitos (National Institute for Health and Welfare, Helsinki, Finland)

U . . . . . Unknown (aetiology of pneumonia)

UK . . . . . United Kingdom

USA . . . . . United States of America

UTI . . . . . Urinary tract infection

VT . . . . . Vaccine-type (in this context, serotypes of pneumococci contained within 10-valent PCV: 1, 4, 5, 6A, 7F, 9V, 14, 18C, 19F, 23F)

UK . . . . . United Kingdom

Unicef . . . . . United Nations Children's Fund

WHO . . . . . World Health Organisation

WTCHG . . . . . Wellcome Trust Centre for Human Genetics (University of Oxford, UK)

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

— T S Eliot, *Little Gidding*, 1942

1

Introduction: The diagnosis of pneumonia and pneumococcal disease in children

1.1 Pneumonia in children

Pneumonia is the greatest cause of mortality in children beyond the neonatal period, accounting for 14% of all deaths in children <5 years of age (young children)[1]. In the South Asian countries of Bangladesh, Bhutan, India, Nepal and Pakistan approximately 1.8 million young children died in 2016, nearly one third of the global number of deaths in young children, with 28% of these deaths attributed to lower respiratory tract infection[7]. Considerable progress has been made since the year 2000, with reductions in pneumonia-specific mortality from approximately 1.7 million deaths in young children in the late 1990s to 921 000 by 2015[1, 8, 9]. This occurred despite an increase the global population of young children from approximately 625 million to 675 million – mainly in countries with

a high burden of pneumonia[10]. Meningitis and sepsis caused far smaller estimated burdens of mortality in South Asia and elsewhere.

Much of this progress has resulted from the transition of low income countries to middle incomes, particularly in the populous countries of Asia, with associated improved maternal and child health. Increased immunisation against pathogens that cause – or predispose children to – pneumonia, including with measles vaccine through the Extended Programme on Immunisations (EPI) schedule supported by the World Health Organisation (WHO), and protein-polysaccharide conjugate vaccines against *Haemophilus influenzae* type b (Hib) and *Streptococcus pneumoniae* (pneumococci), supported by the Gavi Alliance, has also been associated with reduced mortality from pneumonia. Indeed, vaccination with pneumococcal conjugate vaccines (PCV) has averted approximately 240 000 deaths in children <5 years of age since 2011, mainly due to protection from pneumococcal pneumonia, despite only partial uptake by high burden countries (Figure 1.1)[11, 12]. The earlier introduction of Hib vaccination into high burden countries has averted approximately 1.2 million deaths in children <5 years of age since 2000, mainly due to protection from pneumonia associated with Hib infection[11]. The use of antiretroviral therapies for children with HIV has been associated with large decreases in invasive pneumococcal disease[13]. It is probable that this also realises reductions in pneumococcal and all-cause childhood pneumonia[14] in populations with high prevalence of HIV.

The inception of the Millennium Development Goals (MDGs) in the year 2000, including MDG4, which targeted an ambitious two-thirds reduction in the mortality rate of young children by 2015, provided a focused target for public health care in low income settings. The founding of the Gavi Alliance (again, in 2000), and the increasing role of large non-governmental organisations, provided support for efforts to meet this target[10]. Future progress on reducing the mortality and morbidity due to childhood pneumonia will result from improving coverage of existing proven interventions (such as provision of vaccination against measles, pneumococci and Hib), provision of simple effective treatment for pneumonia[15], and improvement of the environment in which young children live[16].

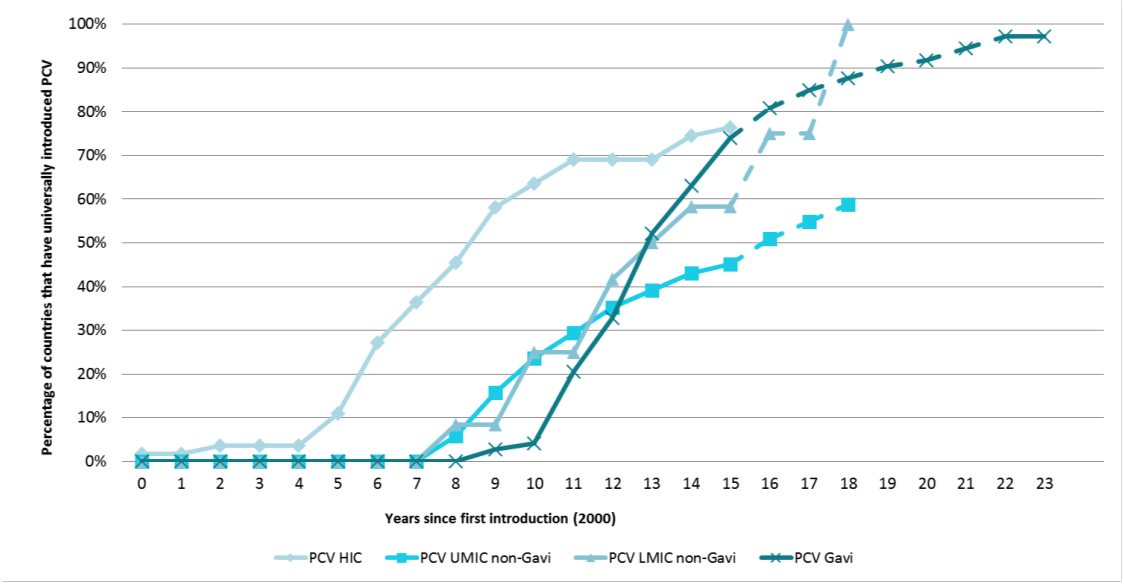


Figure 1.1: Historical pneumococcal conjugate vaccine (PCV) introduction by income group. HIC, high income countries; UMIC, upper middle income countries; LMIC, lower middle income countries; Gavi, eligible for Gavi Alliance support for vaccine introduction. Source: International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health, Vaccine Information Management System, Global Vaccine Introduction Report, December 2015. Used under a Creative Commons Licence.

Prioritisation of interventions also requires an improved understanding of the aetiology of childhood pneumonia. This was the rationale behind the recently concluded *Pneumonia Etiology Research for Child Health* (PERCH) project[10, 17], which assessed the aetiology of severe pneumonia in children aged 1 month to <5 years of age over 8 sites in 7 low and middle income countries (Mali, Gambia, Kenya, Zambia, South Africa, Bangladesh, and two sites in Thailand). This study is described in detail below. Given the complexity of this challenge, it is likely that the findings will emphasise the need for new diagnostic techniques for the aetiology of childhood pneumonia.

Within this context, here I first introduce the clinical features of pneumonia, sepsis and meningitis (all of which may be associated with invasive pneumococcal disease) and describe current state-of-the-art diagnostic tests for the aetiology of childhood pneumonia, with a specific focus on pneumococcal pneumonia. I then describe epidemiological techniques to determine pneumonia aetiology, including the use of data derived from vaccine studies to estimate the burden of disease. Following this I summarise the background to two novel diagnostic approaches for the aetiology of childhood pneumonia: using

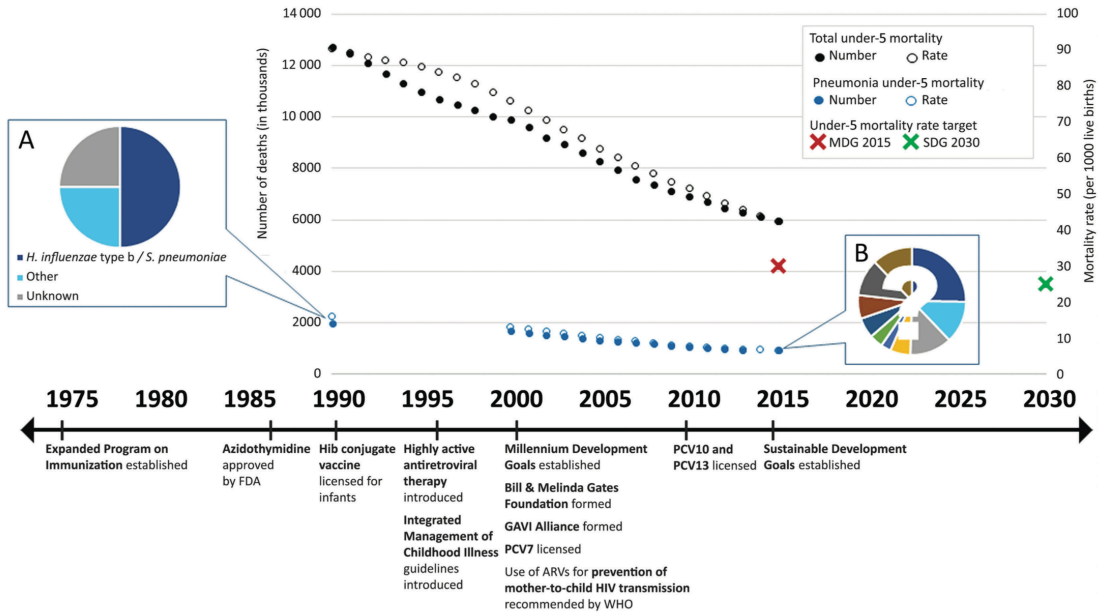


Figure 1.2: Annual <5 years of age all-cause and pneumonia mortality rates with global targets and pivotal health interventions, by time. A: The conceptual pneumonia mortality aetiology distribution from the preconjugate vaccine era, and B: The unknown aetiologic distribution of the present and future. ARVs, antiretroviral drugs; FDA, US Food and Drug Admininstration; Hib, *Haemophilus influenzae* type b; MDG, Millenium Development Goal; PCV, pneumococcal conjugate vaccine; SDG, sustainable development goal. Source: O'Brien et al. 2017[10]. Used under a Creative Commons Licence.

antigen-specific antibody-secreting cells (ASCs), and studies of host RNA expression (transcriptomics).

1.1.1 Clinical features of pneumonia, meningitis and sepsis in children

Pneumonia

The introduction of *Integrated Management of Childhood Illness (IMCI)* guidelines by the WHO and the United Nations Children's Fund (Unicef) in 1995 provided a standardised approach to the management of childhood illness[18]. These guidelines described the assessment and classification of the sick child aged 2 months to <5 years of age, or the sick newborn aged 1 week to <2 months of age, at presentation to healthcare facilities. There was a specific focus on pneumonia, diarrhoea, fever (including measles),

malnutrition and anaemia and immunisation status due to their perceived importance in acute disease, chronic disease, and disease prevention – particularly in low income, high disease burden, settings.

The *IMCI* classification of the sick child was primarily developed as a sensitive but non-specific tool to guide administration of antibiotics in primary care in low income settings. Thus, using the original guidelines, a young child with cough and/or difficulty breathing and "fast breathing" was diagnosed with "pneumonia" and prescribed oral co-trimoxazole; if the child also had chest indrawing or "danger signs" the child was diagnosed as "severe pneumonia" and referred to healthcare facilities for parenteral penicillin therapy (see Table 1.1). These guidelines have recently been modified with fast breathing and/or chest indrawing and classified as "pneumonia" and treated with oral amoxicillin, and only children with "danger signs" (not able to drink, persistent vomiting, convulsions, lethargy on unconscious, stridor when calm or severe malnutrition) classified as "severe pneumonia or very severe disease" with consequential referral to hospital[15]. This reflects evidence that oral amoxicillin given at home is equivalent to parenteral penicillin at treating pneumonia with chest indrawing in low income settings[19], particularly when given at high doses[20]. The modified guidelines instruct parenteral penicillin and gentamicin for pneumonia with danger signs (severe pneumonia or very severe disease). In this thesis I term the diagnosis of childhood pneumonia according to *IMCI* guidelines "WHO clinical criteria for the diagnosis of pneumonia".

Oxygen therapy was presumed to be unavailable at primary care in low income settings, hence *IMCI* guidelines omit measurements of hypoxaemia as part of the clinical diagnosis and classification of pneumonia. However, an increasing recognition of the importance of oxygen therapy for the treatment of childhood pneumonia in low income settings[21, 22] has led many countries to implement a further classification of childhood pneumonia, "hypoxaemic pneumonia" with the aim of providing targeted therapy to these children[23].

The emphasis on sensitivity of the guidelines to pneumonia that could be caused by bacterial infection, over specificity, theoretically ensures that all children who are in

need of antibiotic therapy receive it (even if a relatively large number of children with non-bacterial pneumonia are also treated). This classification scheme also accepts that a large number of children with "pneumonia" (according to WHO clinical criteria) will not have pneumonia, but an alternative diagnosis, including non-infectious respiratory disease, or other infections (that manifest as tachypnoea and fever). The WHO clinical classification was also developed for children 2 months to <5 years of age. Extension to older age groups is problematic, since it requires standardised thresholds for respiratory rate by age or additional criteria, which WHO have not clearly defined in children 5 years of age and older[24].

While acknowledging these limitations, many previous studies of pneumonia incidence and pneumonia aetiology have used the WHO clinical criteria for the diagnosis of pneumonia, restricting enrolment to children <5 years of age. This facilitates comparability of data between study sites, and over time, and limits enrolment to young children, where the greatest burden of mortality from pneumonia has been described[25, 26]. The PERCH study, for example, used the WHO clinical criteria for the diagnosis of pneumonia, but extended the age range to 1 month to <5 years, and excluded children whose clinical features responded to a trial of inhaled bronchodilator therapy[25]. Alternatives to using the WHO clinical classification of pneumonia for study enrolment include using locally-adapted clinical diagnostic criteria[27], or clinician-defined pneumonia[4, 28, 29]. These local criteria may be more accurate for the diagnosis of pneumonia in an individual patient and across a case series of patients, but may also inhibit accurate comparisons between settings. I describe the case definitions used for the studies contained within this thesis in detail in Chapter 2.

The WHO methodology for the standardised interpretation of paediatric chest radiographs was developed as an alternative to clinical criteria for the diagnosis of pneumonia[30] (see Section 1.2.2). In particular, radiographic endpoint consolidation was perceived to be highly associated with bacterial pneumonia, hence its use as an "endpoint" against which to assess the efficacy of PCV in randomised controlled trials in the USA[31], South Africa[32], Gambia[33] and the Phillippines[34]. The association between endpoint consolidation and bacterial pneumonia (particularly pneumococcal pneumonia)

was confirmed by "vaccine-probe" studies that examined the decrease in incidence of endpoint consolidation in children vaccinated in randomised controlled trials of PCV vs unvaccinated children[3].

The WHO methodology for interpreting chest radiographs was originally defined in children <5 years of age. However, it is reasonable to extend these criteria to children 5 to 14 years of age. Children with endpoint consolidation therefore form an important sub-group of all children with pneumonia in the context of epidemiological studies of pneumonia aetiology[35].

Meningitis and sepsis

The *IMCI* guidelines have been extended to include the recognition, classification and treatment of meningitis and sepsis in hospital settings in the *Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illness*[36]. Both meningitis and sepsis are important syndromes associated with pneumococcal infection. Table 1.1 describes the clinical features associated with "suspected meningitis" and "suspected sepsis". As with the diagnosis of childhood pneumonia, sensitivity is emphasised over specificity to prioritise antibiotic treatment for children with possible bacterial infection.

Within this thesis I use "suspected meningitis" to indicate a child who has a suspected infection and one or more clinical features of meningitis on presentation, as described in Table 1.1, such that the child was treated with intravenous antibiotics. However, in the context of vaccine-preventable disease surveillance, more specific case definitions have been employed for the diagnosis of meningitis – on the basis that children with suspected meningitis require hospitalisation and sampling of cerebrospinal fluid (CSF; in the absence of contraindications to lumbar puncture). "Probable bacterial meningitis" is diagnosed as a suspected case of meningitis with CSF that is either turbid, has a leucocyte count >100 cells/mm^3, or has leucocytes of 10–100 mm^3 and either an elevated protein concentration (>100 mg/dl) or decreased glucose concentration (<40 mg/dl)[37]. "Confirmed bacterial meningitis" cases are those with a pathogen identified by

Table 1.1: Clinical features of pneumonia, meningitis and sepsis as determined by the WHO *IMCI* guidelines[18] and *Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illness*[36].

Pneumonia	**Severe pneumonia**[a]	**Pneumonia with danger signs**	**Meningitis**	**Sepsis**
Cough or difficulty breathing with:	Cough or difficulty breathing with:	Not able to drink	History: fever with,	Fever with no focus of infection
Fast breathing:	Chest indrawing	Persistent vomiting	Convulsions	No specific features of meningitis
≥50 breaths/min in a child aged 2–11 months		Convulsions	Vomiting	Confusion
≥40 breaths/min in a child aged 1–5 years		Lethargic or unconscious	Inability to drink	Signs of systemic upset
		Stridor when calm	Head or neck ache	
		Severe malnutrition	Recent head injury	
			Examination:	
			Altered consciousness	
			Neck stiffness	
			Bulging fontanelle	
			Non-blanching rash	
			Lethargy or irritability	
			Evidence of head trauma	
			Signs of raised ICP	
			Decreased consciousness	
			Unequal pupils	
			Rigid posture or posturing	
			Focal paralysis	
			Irregular breathing	

[a]Pneumonia and severe pneumonia have more recently been combined into one category, "pneumonia"[15]. For clarity I refer to them as pneumonia with fast breathing, and pneumonia with chest indrawing in this thesis.

microscopy or culture, or antigen detection methods; more recently this has also included CSF samples with a pathogen detected by polymerase chain reaction (PCR).

Varying definitions of sepsis exist, most recently as the *Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*[38]. In *Sepsis-3*, sepsis is defined as "life-threatening organ dysfunction caused by a dysregulated host response to infection". This is defined on a operational level by changes in arterial blood oxygen concentration (respiratory function), blood platelet count (coagulation function), plasma bilirubin concentration (liver function), mean arterial pressure (cardiovascular function), consciousness level (according to the Glasgow Coma Score), and creatinine concentration and urine output (renal function). A much more sensitive (and less specific) definition of suspected sepsis in children is described in Table 1.1. These clinical features form a practical guide to which children should be admitted to hospital, undergo further diagnostic testing (including culture of bacterial pathogens from blood where available), and start antibiotic therapy. In the context of vaccine-preventable disease surveillance, the definition of sepsis is far more stringently defined as an, "invasive infection which can be diagnosed with microbiological tests on cerebrospinal fluid or blood." Within this thesis I use the terms "suspected sepsis" to indicate a child who has a suspected infection and one or more clinical features of sepsis on presentation, as described in Table 1.1, such that the child was treated with intravenous antibiotics for suspected sepsis. I use the term invasive bacterial disease to indicate a child from whom a bacterial pathogen has been isolated and cultured from a normally sterile site (typically blood, CSF or pleural fluid, and specifically excluding pathogens cultured from urine).

1.2 Pneumococcal disease in children

1.2.1 *Streptococcus pneumoniae* as bacterium, coloniser and pathogen

Pneumococci are Gram-positive organisms that typically form diplococci – although more rarely single cocci or short chains are noted. Pneumococci were first identified in the late 19th century as a major cause of human pneumonia[39]. When grown on agar

containing sheep (or horse) blood, and preferably with a 5% concentration of ambient CO_2, pneumococci can be identified by a surrounding green zone of α-haemolysis and sensitivity to optochin (ethylhydrocuprein hydrochloride). An alternative identification test to the optochin sensitivity test is to assess for solubility of the organisms in bile (sodium deoxycholate; pneumococci are bile-soluble)[40]. Other α-haemolytic pneumococci include the viridans group, usually commensal bacteria of the oropharynx not associated with disease, which may also be non-haemolytic.

In contrast, β-haemolytic streptococci produce clear zones of haemolysis when cultured on blood agar, and are classified according to the presence of Lancefield Group polysaccharides on the cell wall. Group A β-haemolytic streptococci are associated with pharyngitis, sepsis and rheumatic fever, as well as pneumonia, whilst group B streptococci are associated with sepsis in neonates and the elderly. Groups C and G less commonly cause disease.

Pneumococci are routinely classified on the basis of the polysaccharide capsule that surrounds the organism[41]. This capsule, approximately 200–400 nm in thickness, is bound covalently to the outer surface of the cell wall peptidoglycan, and inhibits phagocytosis of the bacterium by macrophages in the absence of anti-capsular antibodies. At least 92 capsular types (serotypes) have been described, with an exceptional degree of polymorphism at the *cps* locus which encodes the enzymes that construct the repeating unit of most polysaccharide capsules[42]. The polysaccharides associated with serotype 3 and serotype 37 pneumococci are an exception, synthesised by genes in the synthase pathway (outwith the *cps* locus) and with a more basic structure[43]. Pneumococci are identified according to their serotype using the Quellung reaction: the stepwise addition of group- and then type-specific sera derived from rabbits inoculated with known serotypes of pneumococci, with the appearance of capsular swelling and a sharply demarcated halo around the diplococci following application of the corresponding serotype-specific serum[40]. Latex agglutination is an alternative serological approach to the serotype-specific identification of pneumococci, with type-specific sera bound to microscopic latex beads[44].

Molecular techniques have, to some extent, superseded culture and serological methods for the detection of pneumococci. Quantitative (real-time) polymerase chain reaction (qPCR) amplification for detection of the pneumococcal genes pneumolysin (*ply*), autolysin (*lytA*) and pneumococcal surface adhesion (*psaA*) genes have been trialled, with *lytA* the most discriminatory vs other streptococcal species[45]. However, the high sensitivity of PCR for *lytA* limits its utility as a diagnostic test from blood since nasopharyngeal (NP) carriage of pneumococci is associated with false positive results (see below[46, 47]). The more tightly regulated environment of the CSF – and possibly a higher pre-test probability of pneumococcal infection in children with meningitis – improves the positive predictive value of *lytA* PCR for pneumococcal meningitis. Serotyping of pneumococci using molecular methods has the potential to increase sensitivity over culture-based methods, particularly in the context of prior antibiotic therapy where culture-based techniques are unlikely to be able to isolate pneumococci, and reduce the cost and labour associated with serological techniques. A large number of alternative molecular serotyping techniques have been developed and have been compared in detail by the PneuCarriage project[48]. The most accurate of these methods (in comparison to rigorous serological techniques) used the BμG@S microarray and bioinformatic techniques for the identification of genes specific to pneumococci. This technique has also provided the ability to examine NP carriage of multiple serotypes of pneumococci at the same time[49].

Nasopharyngeal (NP) carriage of pneumococci is presumed to be a biological prerequisite for pneumococcal disease – whether pneumonia, meningitis, or sepsis, or the milder condition of otitis media[50]. However, prior to the introduction of PCVs the prevalence of pneumococcal carriage was approximately 50–80% in children <5 years of age across a range of settings in urban and rural Nepal[49, 51, 52] and other low income settings in Asia[53–55]. More generally NP carriage of pneumococci was approximately 60% in low income countries and approximately 50% in low middle income countries in children <5 years of age[56]. Yet, the vast majority of children with NP carriage of pneumococci clear the pathogen without developing disease. NP carriage is therefore necessary, but not sufficient for disease to occur. NP carriage may provide a niche for the replication and

transmission of pneumococci between individuals, and potentiates recombination and translocation of genes between pneumococci. In particular, intra-host evolution is driven by the selective pressures of antibiotic therapy[57] and PCVs[58, 59].

Infection with HIV[60–62], malnutrition[29], and a number of inherited immunodeficiencies[63, 64] predispose children to invasive pneumococcal disease. However, given that most infections occur in children who were apparently previously healthy, the expression of virulence factors associated with pneumococci must also be important. Pneumococcal capsular polysaccharides inhibit phagocytosis, and different serotypes have different odds of being associated with invasive pneumococcal disease – which appear to be stable over time and place[65–67]. In particular, serotype 1 and serotype 5 are rarely isolated from NP specimens and yet are highly prevalent amongst pneumococci isolated from childhood invasive disease in Kathmandu in Nepal, other South Asian countries and elsewhere[4, 65, 68, 69], suggesting that they are highly virulent. The mechanistic basis for the high virulence associated with serotypes 1 and 5 has not yet been determined, and may relate either to the capsular polysaccharides themselves, or to associated protein virulence factors. Additional polysaccharides shared between virulent and avirulent serotypes of pneumococci are the cell wall polysaccharide (CWPS; techoic acid) and F-antigen (lipotechoic acid with a similar structure to that of CWPS).

Of protein virulence factors, pneumococcal surface protein A (PspA) is bound to the cell surface through the cell wall polysaccharide and acts to block the binding of complement factor C3, thereby inhibiting opsonisation and phagocytocysis of the bacterium[70]. Choline binding protein A (CbpA) mediates adherence to host epithelial cells, and binds IgA and complement factor H (a key regulator of the alternative pathway of complement activation[71]). Autolysin (LytA) is a surface-bound enzyme that is activated by the host immune system, acting to degrade the pneumococcal cell wall and release the cytoplasmic enzyme pneumolysin (Ply). In turn, Ply binds to cholesterol moieties in host cell walls, introducing transmembrane pores and lysing host immune cells[72]. Additionally, pneumococcal IgA protease is secreted to bind IgA1 (the dominant Ig isotype in the human airway) to inhibit opsonisation of the bacterium[73].

1.2.2 Diagnostic methods for childhood pneumococcal pneumonia

Knowledge of the pathogen-specific burden of pneumonia is essential for the prioritisation of public healthcare interventions, including the introduction of vaccines to routine immunisation schedules. In addition, pathogen-specific diagnostic tests for bacterial pneumonia may be able to enhance the accurate use of antibiotics in individual patients, particularly if antibiotic sensitivity can be detected. Given the global importance of pneumococcal infection, and the focus of this thesis, I concentrate this section on the diagnosis of pneumococcal pneumonia.

Culture of blood (and where possible, pleural fluid) remains the most specific widely used test for the aetiology of bacterial pneumonia[74]. Across a number of settings, prior to the introduction of immunisation schedules containing PCV, the pneumococcus was the most prevalent bacterium isolated from the blood of children with pneumonia[27, 75, 76], and one of the most prevalent causes of any bacteraemia[77–79]. However, the sensitivity of blood culture for bacterial pneumonia is limited by restriction of bacteria to the lung mucosa, pre-treatment with antibiotics prior to sampling, and the small volumes of blood typically obtained from children for culture. Using the "vaccine-probe" approach[80] (detailed below), O'Brien et al.[3] estimated that approximately 13 cases of pneumonia with endpoint consolidation were averted for every 1.9 cases of invasive pneumococcal disease in a trial of a 9-valent PCV in Gambia[33], giving a sensitivity of approximately 15% in this setting (with low prior antibiotic therapy and a high standard of blood culture technique). In others settings with a higher prevalence of prior antibiotic therapy, smaller blood volumes[81], or less efficient culture techniques, sensitivity is likely to be lower.

There is therefore an urgent need for the development of new diagnostic tests for the aetiology of childhood pneumonia that are sensitive and specific. The PERCH study is the largest study of childhood pneumonia ever undertaken and has recently started publishing preliminary data. The goal of the PERCH study was to characterise the aetiology of serious childhood pneumonia in sub-Saharan Africa and South East Asia[10]. To achieve this, the study assessed the aetiology of severe pneumonia in children aged 1

month to <5 years of age over 8 sites in low and middle income 7 countries (Mali, Gambia, Kenya, Zambia, South Africa, Bangladesh, and two sites in Thailand). Over 4000 cases of severe pneumonia were prospectively enrolled, with over 5000 prospectively enrolled age-matched community controls. Current state-of-the-art microbiological techniques were used with results integrated into a Bayesian model to predict the aetiology of pneumonia in each individual case, and across the series of cases.

Estimates of the utility of alternative diagnostic tests for the aetiological diagnosis of childhood pneumonia have recently been published from the PERCH study[10], and are summarised in Table 1.2. These tests were targeted towards direct identification of both bacterial and viral pathogens, and did not use techniques based on the host immune response. I discuss these tests by the site of tissue sampling, and contextualise the results with other data where available.

Samples from the upper airway and sputum

Data from the PERCH study have shown that culture of pneumococci from nasopharyngeal specimens[83] or induced sputum[85, 86] is unable to discriminate between confirmed cases of childhood pneumococcal pneumonia and carriage in children in the community. In contrast, use of qPCR of *lytA* to quantify pneumococci from NP specimens showed significantly higher densities in children with definite pneumococcal pneumonia, endpoint consolidation and high CRP concentration; suggesting that this method might be useful for epidemiological studies of pneumonia aetiology[83], although a sensitivity of 0.64 makes it suboptimal for guiding antibiotic therapy for individual cases. Data from qPCR methods for other bacterial pathogens showed similar findings (Table 1.2[84]).

Quantitation of viruses using qPCR from NP samples or induced sputum yielded significant odds for increasing viral density in children with pneumonia (cases) vs children without pneumonia (controls) for RSV, parainfluenzae 1, parainfluenzae 3, adenovirus, human metapneumovirus, rhinovirus or human bocavirus (Table 1.2[82]). For these viruses, the overlap in densities between cases and controls reduced the ability of viral density for the aetiological diagnosis of a single case of pneumonia, although the

Table 1.2: Summary of diagnostic tests used in the PERCH study for the aetiological diagnosis of childhood pneumonia.

Pathogen(s)	Diagnostic method	Key findings	Limitations	Ref.
Respiratory viruses	qPCR of viral genes from NP specimens	Viral load associated with pneumonia vs non-pneumonia controls for: RSV (OR 2.0), parainfluenzae 3 (OR 1.5), parainfluenzae 3 (OR 1.4), adenovirus (OR 1.3), HMPV A/B (OR 1.2), rhinovirus (OR 1.2), HBOV (OR 1.1).	May discriminate between definite pneumococcal pneumonia and controls, but low specificity for pneumonia at optimal threshold. RSV is a possible exception, where may be adapted for epidemiological studies.	[82]
S. pneumoniae	qPCR of *lytA* from NP specimens	Increased density of NP carriage with *S. pneumoniae* associated with definite pneumococcal pneumonia and EPC, WHO danger signs and raised CRP in non-confirmed cases, in comparison to non-pneumonia controls. Optimal threshold: sensitivity = 0.64 and specificity = 0.92.	May discriminate between definite pneumococcal pneumonia and controls, but low sensitivity at optimal threshold. May be adapted for epidemiological studies.	[83]
H. influenzae, M. catarrhalis, S. aureus	qPCR of bacterial genes from NP specimens	Higher densities of *H. influenzae*, *M. catarrhalis*, *S. aureus* in children with definite Hib pneumonia, and children with pneumonia and EPC.	Low sensitivity and specificity – may be useful for epidemiological studies.	[84]
Bacterial pneumonia	Culture and microscopy of bacteria from induced sputum specimens	No associations between potential pathogen detection and definite bacterial pneumonia or proxy markers of pneumonia.	Not useful in this setting.	[85]
Bacterial pneumonia	qPCR of bacterial genes from induced sputum and NP specimens	Yield of organisms using PCR method on induced sputum only marginally increased in comparison to NP specimens.	Induced sputum provides little additional diagnostic information in comparison to NP specimens.	[86]
Bacterial pneumonia (vs RSV pneumonia)	Plasma CRP concentration	Increased CRP concentrations associated with bacterial pneumonia (*S. pneumoniae*, *H. influenzae* and other bacteria) vs RSV pneumonia. Increased CRP associated with *S. pneumoniae* and *H. influenzae* vs other bacteria. Optimal threshold (CRP 37.1 mg/l): sensitivity 0.77, specificity 0.82.	Not pathogen specific – may be useful for determining antibiotic treatment and epidemiological studies.	[87]
S. pneumoniae	qPCR of *lytA* from whole blood	Increased proportion positive in children with definite pneumococcal pneumonia in comparison to non-pneumonia controls (0.64 vs 0.06).	Low sensitivity and specificity.	[46]
S. pneumoniae	qPCR of *lytA* from whole blood	Increased load of *S. pneumoniae* associated with definite pneumococcal pneumonia and proxy markers of pneumococcal pneumonia. Considerable overlap with definite other bacterial pneumonia and non-pneumonia controls.	May discriminate between definite pneumococcal pneumonia and controls, but low sensitivity at optimal threshold.	[47]

Respiratory viruses: Adenovirus, coronavirus 229, coronavirus 43, coronavirus 63, coronavirus HKU, influenza A, influenza B, influenza C, HBOV, HMPV A/B, parainfluenzae 1, parainfluenzae 2, parainfluenzae 3, parainfluenzae 4, PV/EV, rhinovirus, RSV.
Abbreviations: qPCR; quantitative PCR; EPC, radiographic endpoint consolidation; HBOV, human bocavirus; HMPV, human metapneumovirus; PV/EV, parechovirus/enterovirus; RSV, respiratory syncytial virus; OR, odds ratio per 1 $\log_{10}$ increase in copies/ml; Hib, *Haemophilus influenzae* type b

data could be usefully incorporated into a predictive model to assess the aetiology-specific burden of childhood pneumonia in an epidemiological context[47]. In summary, microbiological and molecular analysis of NP samples and induced sputum were not able to identify the aetiology of pneumonia in individual cases in the PERCH study, and had limited ability to differentiate between grouped cases of pneumonia and non-pneumonia controls.

Pneumococcal serotype-specific odds ratios for NP carriage in pneumonia vs carriage in children in the community have not yet been published by the PERCH study. As detailed above, a meta-analysis of global serotypes[65], and data from the UK[66] and Israel[67], have shown relatively stable odds ratios for prevalence of pneumococcal serotypes amongst invasive pneumococcal disease isolates vs NP carriage prevalence. The serotype-specific "disease potential" can be estimated as an odds ratio ($\text{odds}_{pneumonia}/\text{odds}_{community}$) with appropriate adjustments for confounders[67]. Thus, NP carriage of a pneumococcal serotype with a "high" disease potential may be (imperfectly) predictive of pneumococcal pneumonia in an individual patient.

Urine sampling

Testing of urine for the presence of pathogen antigens from either pneumococci or *Legionella pneumophila* is used widely for the diagnosis of these two pathogens in adults with pneumonia, with a sensitivity of 70–80% and a specificity of >90% for pneumococcal disease. In healthy adults, there is a low prevalence of NP carriage of pneumococci. Unfortunately the BinaxNOW pneumococcal test (based on detection of pneumococcal cell wall polysaccharide) detects pneumococcal antigens from children with healthy NP carriage in addition to those from children with disease, making it unhelpful to discriminate between potential pneumonia aetiologies[88, 89]. Urinalysis for pathogen-specific metabolites ("metabolomics") is currently being investigated as a potential diagnostic strategy. The BinaxNOW pneumococcal test has been used for the diagnosis of pneumococcal infection from samples of CSF and pleural fluid, where it is likely to be highly accurate with estimated sensitivity and specificity of 95–100%[90, 91].

Blood samples

Molecular methods to detect pathogen nucleic acid and serological testing have been used for the aetiological diagnosis of pneumonia in children, in addition to isolation and culture of bacterial pathogens from blood described above. Prior to the PERCH study data, limited data using qPCR for *lytA* from the blood of healthy children in Italy with NP carriage of pneumococci suggested that these methods might be specific for pneumococcal disease[92]. Further, data from samples from children with known invasive pneumococcal disease[92] or with pneumonia[93] in Italy suggested that qPCR for *lytA* was also sensitive to invasive pneumococcal disease, and associated with "proxies" for pneumococcal pneumonia (such as pneumonia with high plasma CRP concentrations). The authors[93] were able to serotype pneumococci using the *cps* gene in the majority of qPCR positive cases.

The PERCH study used a multiplexed qPCR for the detection and quantitation of a number of viral and bacterial pathogens, including pneumococci, from blood (Table 1.2[2, 46, 47]). Here, using data from children in largely low income countries, increased pneumococcal density was associated with definite pneumococcal pneumonia vs healthy controls[46]. In the PERCH study, pneumococcal DNA was detected in all children with definite pneumococcal pneumonia. Using an optimal threshold for pneumococcal density to discriminate between microbiologically-confirmed pneumococcal pneumonia and healthy controls, 62.5% of children with confirmed pneumococcal pneumonia were positive as were 9.3% of cases with another confirmed (non-pneumococcal) pathogen. Additionally 4.3% of cases with no other confirmed pathogen, and 3.1% of healthy controls were also positive[47]. Thus, the large dataset collected by the PERCH study suggests that although qPCR for *lytA* may be useful predicting pneumococcal aetiology in epidemiological studies, sensitivity and specificity are too low for it to be recommended for routine clinical care.

The use of serological testing for the diagnosis of pneumococcal pneumonia has been reassessed following the development of assays to assess antigen-specific pneumococcal antibody levels. These assays were primarily developed to assess vaccine-induced immune responses, but have been adapted for pneumonia aetiology studies[94–96].

IgG responses to pneumococcal capsular polysaccharides and proteins have been demonstrated in young infants associated with NP carriage of pneumococci and provide some protection against pneumococcal carriage, pneumonia and invasive pneumococcal disease[97–102]). The presence of pneumococcal antibodies in healthy children means, however, that serological diagnosis of pneumococcal infection requires measurement of acute and convalescent antibody concentrations.

Serology has been used prospectively for the diagnosis of pneumococcal pneumonia in Finnish children using a threshold of a two-fold rise in antibody titres, as measured by enzyme-linked immunoassays (ELISAs) to detect IgG to Ply and C-polysaccharide[103]. In this case series of children with radiographic infiltrates or endpoint consolidation, 37% of children had a positive serological test for Ply and were diagnosed as a case of pneumococcal pneumonia. Using this threshold, approximately 3% of healthy children would have a false positive test to Ply, with <1% of young adults developing a positive test to Ply during an upper respiratory tract viral infection[104, 105].

Further supporting evidence for the use of serology to pneumococcal proteins has come from the use of ELISAs to detect IgG to eight pneumococcal proteins in the serum of 13 Brazilian children with invasive pneumococcal disease and a control group of 23 Finnish children with viral pharyngitis[106]. A 2-fold increase in titres to Ply, CbpA, PspA1, PspA2, pneumococcal choline binding protein A (PcpA), pneumococcal histidine triad protein D (PhtD), serine/threonine protein kinase-C (Stkp-C) and protein for cell wall separation of Group B streptococci (PcsB-N) gave a sensitivity of 23–76% and a specificity of >90% between cases and controls depending on the protein assayed. The clear limitation to this study is that the cases and controls came from populations that differ in significant ways, including PCV immunisation status, geographic region, and socioeconomic status, and serological status may be related to these confounders rather than case vs control status.

In theory, assay of IgG to pneumococcal capsular polysaccharides may both diagnose pneumococcal pneumonia, and inform the development of higher valency PCVs by providing data on the serotype distribution. The recently WHO-validated IgG and IgA

ELISA to 9 pneumococcal capsular polysaccharides was used for the diagnosis of pneumococcal pneumonia in children with confirmed pneumococcal pneumonia (35 children) and suspected or possible pneumococcal pneumonia (128 children) in Belgium, following the introduction of 7-valent PCV[107]. For confirmed pneumococcal pneumonia (of a serotype within the IgG ELISA), the sensitivity was 82% with a high degree of concordance with the Quellung reaction, and there was typically a 10-fold change in antibody concentrations to the serotype detected by culture. The IgG response to invasive pneumococcal disease may be different to that of non-invasive pneumococcal pneumonia ("spectrum bias"). The authors therefore investigated serotype-specific IgG acute–convalescent concentrations in children with suspected pneumococcal pneumonia in children with negative blood cultures, but with pleural effusion and CRP concentration >40 mg/l and white cell count >25 x 10^3/mm^3. If these children with suspected pneumococcal pneumonia, 55% had acute–convalescent rises in serotype-specific IgG similar to those exhibited by children with invasive pneumococcal disease in the study. However, the deliberate choice of samples from children with a high *a priori* probability of pneumococcal pneumonia may inflate the positive predictive value of the test and the absence of a control group makes the negative predictive value incalculable. Similar findings were also found in an unvalidated serotype-specific IgG ELISA in a case series of 196 Italian children 2 to 5 years of age[108].

In addition to pathogen "specific" markers of bacterial infection, inflammatory markers such as plasma/serum C-reactive protein (CRP) concentrations, procalcitonin concentrations, or blood neutrophil count may also be useful to discriminate between viral and bacterial causes of pneumonia[87, 109–112]. CRP concentration is increased in any severe bacterial infection (regardless of specific pathogen), and this is sufficiently sensitive and specific to have been incorporated into routine clinical practice. If bacterial pneumonia is predominantly caused by a single pathogen then the pathogen-specificity of CRP concentration will be high in this setting. For example, in settings without vaccination against pneumococci, but with high prevalence of vaccination against Hib, CRP concentration may be relatively specific for pneumococcal pneumonia. However,

this finding may not be generalisable to other settings with a different pathogen-specific burden of pneumonia – another good example of spectrum bias.

Lung samples

Direct sampling of the lung may be the most accurate method to diagnose the aetiology of pneumonia in an individual case. Yet lung aspirates are rarely performed either for clinical purposes or for pneumonia aetiology studies due to concerns over the safety of the technique. Lung aspiration requires the insertion of a needle through tissues over the top of a rib, avoiding great vessels and other vital structures, with the withdrawal of a small amount of lung tissue for microbiological testing. In the few centres that practise lung aspiration, the technique is therefore limited to peripheral areas of consolidation for safety reasons [113]. Complications of the technique include pneumothorax (in <1% of patients).

Most data on the use of lung aspiration for the aetiological diagnosis of pneumonia has been from the Medical Research Council Unit in Banjul, Gambia. Here, over a period of 25 years data were available on >500 lung aspirates from five separate studies. The proportion of lung aspirates from which a bacterial respiratory pathogen was isolated was 35–56%; contrasted with 7–25% of blood cultures. Specificity was estimated to be near perfect (with a high degree of concordance for bacterial isolated between blood and lung aspirates)[114]. The addition of molecular diagnostic methods (including qPCR to *lytA* and *cps* for pneumococci) to lung aspirates from children increased sensitivity, with pneumococci detected in samples from 91% of children with severe pneumonia by qPCR and 38% of samples by culture. More than one potential pathogen was identified from the same lung aspirate in 53% by qPCR, with bacterial-bacterial co-detection in 40% of cases[115]. Given that PCR primers for bacterial pathogens are highly specific for the detection of corresponding DNA sequences, and that the healthy lung is assumed to be sterile, this suggests that pneumonia is secondary to infection with multiple pathogens in a large number of cases. The attribution of pneumonia to multiple simultaneous pathogens is an important and challenging finding.

1.2.3 Vaccine probes to estimate childhood pneumococcal pneumonia

An alternative approach to assessing the prevalence of pathogen-specific aetiology in childhood pneumonia is afforded by randomised, placebo-controlled, trials of pathogen-specific vaccines – the "vaccine-probe" approach[80, 116]. The basic concept of vaccine-probe approaches is relatively simple, given data from a controlled cohort design study (e.g. a randomised controlled trial). In its simplest (unadjusted) form, the incidence of pneumococcal pneumonia can be estimated by the difference in the incidence of clinical pneumonia (or pneumonia with radiographic endpoint consolidation) between children who were randomised to receive PCV and those who were not.

The fundamental equations are (for pneumococcal pneumonia and PCVs):

$$VE_{pneumococcal} = VE_{pneumonia}/P_{pneumococcal} \tag{1.1}$$

Where $VE_{pneumococcal}$ is the vaccine efficacy against pneumococcal pneumonia (derived from estimates from definite pneumococcal disease – typically invasive pneumococcal disease), $VE_{pneumonia}$ is the vaccine efficacy against pneumonia as measured by a randomised clinical trial, and $P_{pneumococcal}$ is the proportion of pneumonia secondary to pneumococcal infection (unknown). Rearrangement of the equation yields:

$$P_{pneumococcal} = VE_{pneumonia}/VE_{pneumococcal} \tag{1.2}$$

Two key assumptions are made: that the PCV trialled is specific to pneumococci (and is not associated with non-specific benefits), and that vaccine efficacy against pneumococcal pneumonia is similar to that of vaccine efficacy against invasive pneumococcal disease. In addition, $P_{pneumococcal}$ might be better termed "vaccine-preventable disease incidence"[116], since serotypes of pneumococci not included in PCVs will not be prevented. In practice, $VE_{pneumococcal}$ is adjusted using the serotype distribution of isolates from invasive pneumococcal disease from the setting.

Trials of nine-valent PCV (Gambia, [33], and South Africa[32]), 11-valent PCV (The Philippines[34]), and seven-valent PCV (USA[117, 118]), showed a decrease in radiographic endpoint consolidation in vaccinated children versus those given placebo of 25–60%. When adjusted for the efficacy of the vaccine formulation against the local serotype distribution of invasive pneumococcal disease, efficacy of the vaccine against pneumococcal serotypes contained within the vaccine, and the use of conjugate vaccine against *Haemophilus influenzae* type b, the estimated proportion of pneumonia with radiographic endpoint consolidation caused pneumococcal infection across all settings combined was 36%.

Although clearly not applicable to the diagnosis of an individual child, this approach has been used to model the global burden of pneumococcal disease[3, 11], using country-level estimates of the incidence of pneumonia, meningitis and sepsis. In the year 2000, 826 000 (uncertainty range 582 000, 926 000) deaths were attributed to pneumococcal infection in children <5 years of age, from a total of approximately 10.6 million deaths in this age group[3]. The greatest number of these deaths occurred in African countries, although India had the greatest number of deaths of children <5 years of age secondary to pneumococcal infection of any individual country. Social and economic development, improved access to vaccination, improved access to antiretroviral therapies, and an international focus on child mortality through the MDGs, are likely to have significantly reduced the absolute number of deaths in young children attributable to pneumococcal infection[10, 11].

1.3 Novel diagnostic tests for the aetiology of pneumonia

As described above antigen detection and PCR-based tests increase yield of organisms from the nasopharynx, blood, urine and lung. However, the extent to which these organisms are pathogenic in an individual case is difficult to define – hampering accurate diagnoses. An alternative diagnostic algorithm would examine the childhood immune response to the pathogen. Unfortunately serological assays for pneumococcal infection, and most other respiratory pathogens, have limited specificity in the acute phase or

require convalescent samples[107, 108, 119], or cannot delineate acute/recent infections from past infections.

In contrast, direct measurement of the numbers of antigen-specific antibody-secreting cells (ASCs, a subset of B cells) using enzyme-linked immunospot (ELISpot) assays, or measures of their secretions using the test "antibodies from lymphocyte supernatant" (ALS) may be suitable as diagnostic tests for the aetiology of childhood pneumonia. ASCs are not detectable in peripheral blood in children in health using normal methods[120], and therefore their detection during illness may be specific[121]. Characterisation of antibodies actively secreted by ASCs may have the additional benefit of providing data on the importance of different pneumococcal antigens in inducing a humoral immune response, which may be of interest for vaccine development. "Host" transcriptomic techniques have also been used to accurately distinguish aetiology of infection by quantifying the level of expression of each gene across the entire human transcriptome. The "pattern" of gene expression can also be used to derive an analysis of the network of pathways activated by an infection in an individual, or within groups of children with the same infection.

Here, I briefly summarise the ASC response to infection before introducing the methods and background to the use of ALS and ELISpot assays for the aetiological diagnosis of infection. I then describe data from host transcriptomic techniques used to investigate the aetiology of childhood infection.

1.3.1 The ASC response to infection

Overview of the ASC response to infection

B cells achieve remarkable efficiency in the immune response to infection by "transferring" the antigen-specificity of their B cell receptor (BCR; a membrane-bound form of their unique secretory antibody) to effector mechanisms to protect the host from pathogens. Broadly, four mechanisms achieve this: opsonisation of foreign antigen by antibody, interactions between antibody and complement, B cell presentation of foreign antigen to

other effector cells, and B cell regulation of the immune response. Additionally B cells are involved in the formation of germinal centres[121].

Each B cell clone produces a unique antibody that binds to foreign antigen with a level of affinity during infection[122]. Opsonisation of foreign antigen by cognate antibody induces phagocytosis by macrophages, or killing by neutrophils or cytotoxic T cells. In addition, Antibody-antigen complexes may activate the complement cascade through the classical pathway, yielding C3b which opsonises the foreign antigen to upregulate B cell activation or induce macrophage phagocytosis, or induce the lysis of bacteria or virions through the membrane attack complex[123]. In addition, B cells are important antigen-presenting cells. On pathogen entry, naive B cells capture foreign antigen by their cognate BCRs. The antigen is then internalised, processed and presented by major histocompatibility complex II molecules on the B cell membrane to CD4+ T helper (Th) cells, initiating differentiation towards antibody secretion. Finally, B cells are emerging regulators of the immune response to infection and of autoimmunity through expression of specific cytokines[124]. These include the "pro-inflammatory" cytokines, granulocyte macrophage colony stimulating factor[124], interleukin (IL)-17[125], and IL-6[126], and the "anti-inflammatory" cytokines, IL-10 and IL-35[127, 128].

B cells receive activating signals by at least two pathogen non-specific mechanisms: the cytokine context of the immune response, and pattern recognition receptors; and two antigen-specific mechanisms: BCR-antigen interaction, and interactions with cognate T cells via the T cell receptor (TCR) and other ligands. The integration of these signals determines cellular fate[129, 130].

The cytokine context of an immune response (reviewed in detail elsewhere[131]) physiologically balances pro- and anti-inflammatory effectors to clear an infection whilst limiting damage to surrounding tissues. Activation of pattern recognition receptors, such as toll-like receptors (TLRs), by repetitive pathogen motifs such as lipopolysaccharides, induces differentiation of B cells to more mature ASCs[132] in a T cell independent manner. The limited antigen-specificity of this response is balanced by its rapidity, and this mechanism

may be particularly important in the early immune response to a potentially severe infection[133].

Cognate antigen can activate B cells via T cell independent or T cell dependent pathways[130]. T cell independent BCR-antigen signalling does not lead to somatic hypermutation and affinity maturation. In contrast, T cell dependent pathways require the endocytosis and presentation of pathogen fragments on major histocompatibility type II molecules of the B cell. Such B cells, in the presence of the inflammatory cytokines IL-4, IL-5 and IL-21, upregulate expression of interacting proteins, such as CD40, with nearby activated Th cells. Th cells expressing a cognate T cell receptor (TCR) interact with the antigen expressed on the B cell, and other ligands such as CD40L, thus activating the B cell. ASCs, derived from this extrafollicular response initially secrete antibodies with low affinity for the cognate antigen, and may or may not undergo class-switching. And, like ASCs from the T cell independent response, they also transiently migrate from secondary lymphoid organs through the bloodstream. A proportion of ASCs activated by the T cell dependent pathway also re-enter the B cell follicle where, supported by T follicular helper cells[134], they rapidly proliferate to form germinal centres, and initiate class-switching and somatic hypermutation to undergo the process of affinity maturation[122]. The differentiation of B cells to ASCs in response to infection is summarised in Figure 1.3. Recently matured and migratory ASCs are also known as plasmablasts.

1.3.2 Measuring ASCs in infection

Techniques for measuring the kinetics of recently-activated ASCs in response to infection have recently been reviewed in detail[121]. In brief, across almost all bacterial and viral infections for which data are available, density of antigen-specific ASCs increase in peripheral blood between approximately days 4–7 following onset of symptoms, before reaching a plateau and declining from days 10–14 and becoming undetectable in convalescence. There is considerable variation in the peak number of ASCs detected, which may be partly related to host genetic factors. It is also likely that the ASC response to natural infection is dependent on the quantity of antigen to which the host is exposed,

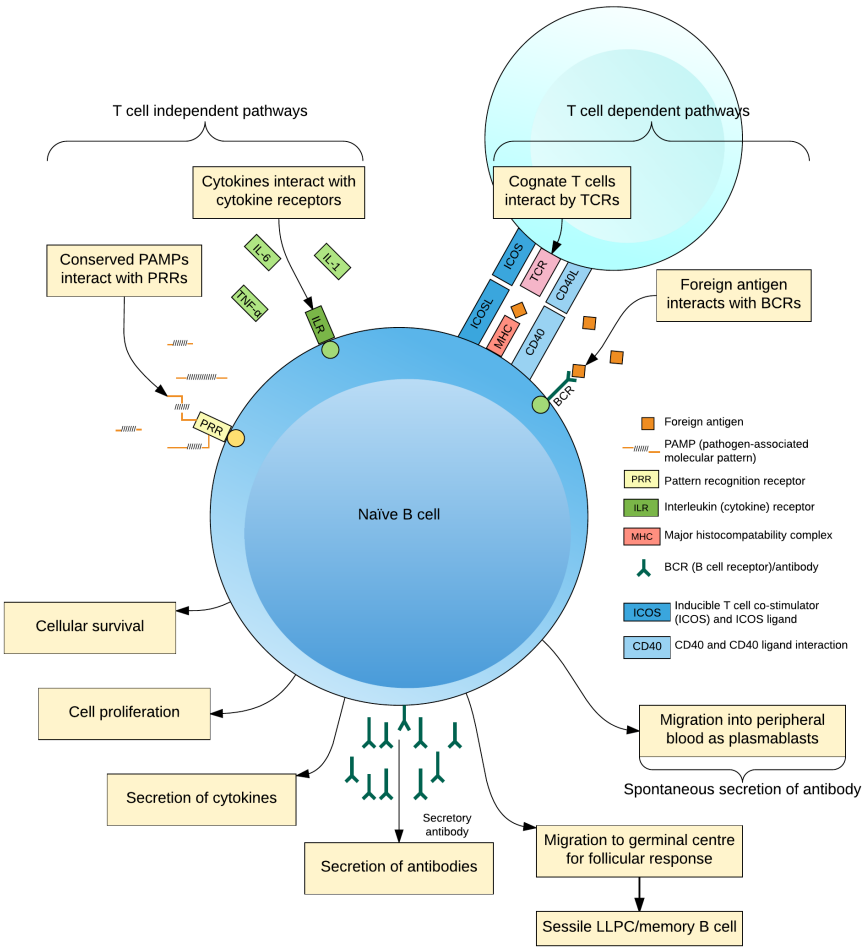


Figure 1.3: Differentiation of B cells in response to antigen. B cells receive activating signals by at least two pathogen non-specific mechanisms, pattern-recognition receptors (PRRs), and the cytokine context of the immune response; and two antigen-specific mechanisms, interactions with cognate T cells via the T cell receptor (TCR) and other ligands and B cell receptors (BCR)-antigen interaction. Integration of these signals predisposes the fate of the maturing B cell to a certain phenotype for the lifespan of the cell, simplistically resulting in plasmablasts, long-lived plasma cells (LLPCs), and memory B cells. Source: Carter et al. (2017)[121]. Used by kind permission of the authors.

and whether this exposure occurs on mucosal surfaces, in secondary lymphoid organs, or in the systemic circulation. Onset of natural infection is typically difficult to define, particularly since mucosal colonisation with a pathogenic organism is likely to be a necessary precursor to many diseases[50, 135, 136]. Practical limitations on sampling blood from patients with infections have also limited the temporal resolution of studies of the ASC response to infection.

Enzyme-linked immunospot assays

Enzyme-linked immunospot (ELISpot) assays typically identify antigen-specific ASCs within a sample of peripheral blood mononuclear cells (PBMCs) by the binding of BCR to antigens and subsequent fluorescence of cells following addition of substrate. Each spot within a well of the ELISpot plate represents a single antigen-specific ASC[137]. ELISpot can be applied to populations of cells that are either sorted by flow cytometry, or more simply PBMCs separated from whole blood using density-dependent centrifugation. Unfortunately, harvesting of cells following density-dependent centrifugation is highly variable, and therefore estimating the number ASCs per unit of blood unreliable. As such results from ELISpot using unsorted ASCs can only be expressed as a proportion of PBMCs. Although simplified protocols for ELISpot exist, it is a difficult technique to optimise, and is reliant upon prespecified antigens to coat the wells of the ELISpot plate.

Assay of ALS

Assay of ALS was developed to combine the antigen specificity of IgG (or IgA) with the temporal specificity of ASCs that transiently circulate during the course of an infection, and the flexibility of a serum sample for the choice of antibody assayed, since ALS can be frozen and transported to centralised laboratories[138]. Assay of ALS requires PBMC separation via density-dependent centrifugation of whole blood, PBMC washing to remove contaminating plasma antibodies, incubation of a pre-defined concentration of PBMC suspension for 24–48 hours, and harvesting of the resulting cell media that is enriched for antibodies that have been actively secreted during the cell culture step. This relatively simple technique has considerable potential to be developed into a diagnostic test for the aetiology of acute infections, and has already been applied with some success

to a number of acute infections including *Escherichia coli* infection, cholera, typhoid, tuberculosis, influenza A and RSV infection, dengue virus infection and mumps virus infection. Table 1.3.2 is a comprehensive list of infections for which assay of ALS and ELISpot assays have been used for the diagnosis of acute infection[121].

Table 1.3: Studies investigating the use of ELISpot and ALS for the diagnosis of infection. Redrawn from Carter et al. 2017[121] by kind permission of the authors.

Disease or pathogen	Setting	Methods and study participants	Results	Ref
Enterotoxigenic *E. coli* dysentery	USA (challenge study)	20 adult volunteers with with live oral attenuated ETEC vaccine; 20 adult volunteers with ETEC oral challenge. Comparison of ELISpot assays for IgA ASCs and IgA ALS (for IgA) at day 0, day 7 and day 21 following challenge. Antigens for ALS time-resolved fluorescence and ELISpot: CFA/II (vaccinees); LT, CS3, CS1 (infected subjects).	ALS against *E. coli* antigens IgA: positive in 75–95% of infected subjects. ELISpot: positive in 85–100% of infected subjects (depending on antigen measured). ALS against CFAII IgA: positive in 75% of vaccinees; ELISpot: positive in 90% of infected subjects. Both ALS and ELISpot considerably higher than faecal IgA measurements against matched antigens.	[139]
	Urban Bangladesh	46 adult and child patients with CS6-expressing ETEC diarrhoea; 10 local adult controls, and 10 local child controls. Comparison of ELISpot assays for ASCs on 12 adult patients, and ALS for all 46 patients, at day 2 and day 7, versus ELISpot and ALS on healthy controls. Antigens for ALS and ELISpot: CS6 and CS5 + CS6.	ALS against CS6: 89% with positive with IgA, 100% positive with IgG. ELISpot against CS6: 100% positive with IgA (GMC 430 ASC/million PBMC), 75% positive with IgG. No responses detectable by ALS or ELISpot in healthy controls.	[140]
***Shigella* spp. dysentery**	USA (challenge study)	30 adult volunteers with oral live attenuated *S. dysenteriae* type 1a vaccine; 10 adult volunteers with S. flexneri 2a oral challenge. Comparison of ELISpot assays for ASCs and ALS (for IgA and IgG). No details of timepoints. Antigens for ELISA of ALS and ELISpot: *Shigella*-LPS.	ALS for for *Shigella*-LPS IgA ASCs: sensitivity of 79%, specificity of 75% versus ELISpot. ALS for for *Shigella*-LPS IgA ASCs: sensitivity of 79%, specificity of 75% versus ELISpot. Details on timing or magnitude of ASC responses by ELISpot of ALS not given.	[141]

Cholera	Urban Bangladesh	30 adult patients with acute cholera and 10 local healthy controls. Comparison of ELISpot assays for IgA ASCs and IgA ALS at presentation to hospital and day 7. Antigens for ALS and ELISpot: CTxB, cholera LPS and MSHA.	Peak cholera-specific ASCs at day 7 (median peak CtxB-specific IgA ASCs ~ 350/million PBMCs, LPS-specific IgA ASCs ~900/million PBMCs). ALS for IgA CtxB-specific ASCs: Sensitivity of 94% versus vibriocidal serum response. ELISpot for IgA LPS-specific ASCs: Sensitivity of 94% versus vibriocidal serum response.	[142]
	Urban Bangladesh	17 children 3–5 years of age, 17 older children and 68 adults with acute cholera. Comparison of vibrocidal antibodies, antigen-specific antibodies, ELISpot for IgM/A/G ASCs and (stimulated) memory B cells at day 2, day 7 and day 30.	Similar antibody/ASC responses across age groups. Returned to baseline by day 30. Peak cholera-specific ASCs at day 7 (median peak CtxB-specific IgA ASCs ~100–200/million PBMCs, LPS-specific IgA ASCs ~200–400/million PBMCs).	[143]
	Urban Bangladesh	9 adult patients with acute cholera infection, 8 adults vaccinated with oral cholera vaccine (Dukoral). Comparison of ASC responses between infection, and post-vaccination using ELISpot (antigens: CtxB and cholera LPS) on days 2, 7 and 30 (infection) and days 0, 7, and 21 (vaccination)	Median peak CtxB-specific IgG ASCs at ~370/million PBMCs (patients) and ~200/million PBMCs (vaccinees). Detectable CtxB/LPS-specific IgA/IgG ASC responses in 9/9 patients and 5/8 post-vaccination at day 7 (returned to baseline during convalescence)	[144]
Enteric fever	Urban Bangladesh	112 adult and child patients with fever of unknown aetiology of 3–7 days duration. Comparison of ALS and plasma antibodies versus comparator standards (classic clinical presentation, Widal titre, blood culture) at day 0, day 5 and day 20 following presentation. Antigens for IgA and IgA ALS: typhoid LPS, whole cell preparation, and membrane preparation.	IgA ALS against LPS and whole cell preparation: sensitivity 88% versus blood culture. IgA ALS and plasma antibodies across comparator groups supports possibility of ALS detecting blood culture negative enteric fever.	[145]
	Rural Bangladesh	243 adult and child patients with clinically-suspected enteric fever (39 blood culture positive), 32 febrile (non-enteric fever) controls (both groups with fever of 3–7 days duration), and 74 local healthy controls.	Sensitivity of 100% for blood culture positive enteric fever, specificity of 78–97% (depending on definition of a true negative case).	[146]

		ALS at presentation to hospital/enrolment (all patients/controls), and day 7 and day 21 in 38 patients. Antigens for IgA ALS: typhoid membrane preparation.		
Tuberculosis	Urban Bangladesh	49 adult patients with smear-positive pulmonary TB; 35 adult patients with other pulmonary diseases; 35 healthy local controls. ALS at presentation to hospital only. Antigens for IgG ALS: BCG and PPD.	IgG ALS against BCG: sensitivity of 93% and specificity of 80% to detect smear-positive in pulmonary TB.	[147]
	Urban Bangladesh	58 child patients with clinical TB; 16 child patients with other infections; 58 healthy child controls. ALS for all patients/controls (and standard microbiology for patients) at day 1, day 60 and day 180 following presentation. Antigens for ALS: BCG.	Of patients with clinical TB: 15% positive by culture of sputum/gastric lavage, 64% positive by published clinical scoring charts, 91% positive by ALS. All children with TB had significantly higher ALS titres that non-TB patients or controls. Declining ALS titres in children with active disease following treatment.	[148]
	Urban Bangladesh	212 adult patients with suspected TB; 100 adult patients with other diseases; 25 healthy adult controls. ALS at presentation to hospital only. Antigens for ALS: BCG and other *M. tuberculosis* antigens.	Sensitivity of 91%, specificity of 88% specificity using BCG-specific IgG ALS for pulmonary TB (smear/culture positive) versus non-TB patients.	[149]
	Urban Ethiopia	84 adult patients with smear negative symptomatic TB (pulmonary, pleural and lymph node); 12 adult patients with other diseases; 45 adult patients with latent TB; 40 non-TB controls. Comparison of IgG ALS, BCG-specific plasmablast enumeration (flow cytometry), Quantiferon Gold, tuberculin skin test. Antigens for IgG ALS: BCG	BCG-specific plasmablasts median of 6% of plasmablasts in symptomatic TB versus 3% in latent TB. ALS: 86–90% sensitivity, 80% specificity for symptomatic TB versus latent TB. ALS: 86–91% sensitivity, 94% specificity for symptomatic TB versus non-TB controls	[150]

	Urban Bangladesh	224 patients <5 years of age with severe malnutrition and radiological features of pneumonia/pulmonary infection. 15 patients with microbiologically-confirmed TB; 41 patients with non-confirmed TB; 168 patients not TB ALS at presentation to hospital. Antigens for IgG ALS: BCG	Similar proportions of "positive", "bordeline" and "negative" ALS results between three groups (confirmed TB, non-confirmed TB, not TB) Sensitivity of 67% and specificity of 51% for confirmed TB versus not TB.	[151]
	Urban China	62 adult patients with culture-positive pulmonary TB; 53 adult patients with other pulmonary diseases; 47 healthy adult controls. IgG ALS at presentation to hospital only. Antigens for IgG ALS: *M. tuberculosis* acid phosphatase.	ALS: 97% sensitivity, 85% specificity for culture positive TB versus healthy controls. ALS: 68% sensitivity and 93% specificity for culture positive TB versus patients with other pulmonary diseases.	[152]
Pneumococcus and other bacterial respiratory pathogens	Finland	16 adult patients with bacteraemic pneumococcal pneumonia and 14 healthy controls. ELISpot to detect pneumococcal-specific IgM/A/G ASCs on day 7 following bacteraemia or from healthy controls. Antigens for ELISpot: patient's pathogen, or mixed capsular polysaccharides (controls).	All patients had pneumococcal-specific IgG ASCs >10/million PBMCs (range 10–1000/million); all controls had 0–1/million PBMCs detected: sensitivity and specificity of 100% in this small retrospective case series.	[153]
	Finland	24 adult patients with acute sinusitis of acute tonsillitis and 9 adult healthy controls. *S. pneumoniae*, *H. influenzae* and α-haemolytic streptococci isolated from patient swabs. Patient blood samples taken 7–14 days following onset of symptoms.	All patients had pathogen-specific IgG ASCs >10/million PBMCs (range 10–1000/million PBMCs); controls had no pathogen-specific ASCs detected: sensitivity and specificity of 100% in this case series.	[154]

		ELISpot to detect antigen-specific IgM/A/G ASCs patient's pathogen, and pathogen panel for controls.		
UTI caused by *E. coli* and other bacteria	Finland	10 adult patients with lower UTI and 17 patients with pyelonephritis due to *E. coli*, *Enterococcus* spp., *Enterobacter* spp. and staphylococci. Patient blood samples 7 days following onset of symptoms and in convalescence. ELISpot to detect antigen-specific IgM/A/G ASCs to patient's own pathogen.	Pathogen-specific IgA ASCs in 100% of patients with pyelonephritis (mean 787, 95% CI 738–836/million PBMCs) and 70% of patients with lower UTI (mean 6 95% CI 6–12/million PBMCs). Lower responses for other Ig isotypes. Downward trend in ASCs at convalescence, but pathogen-specific ASCs still detectable at 7 weeks in some patients.	[155]
	Finland	14 adult patients with lower UTI and 11 patients with pyelonephritis due to *E. coli*. (Patients previously reported in Kantele et al. 1994[155]). Patient blood samples at 7 days following onset of symptoms. ELISpot to detect whole cell, P fimbria or OmpA-specific IgM/A/G ASCs.	IgA ASC responses greater than IgG and IgM; and responses to whole cell greater than to P fimbria and OmpA. Mean P fimbria-specific IgA ASC 294/million PBMCs. P fimbria-specific IgA ASC responses in 100% of patients with P fimbria positive *E. coli* pyelonephritis, and 20% of those with P fimbria lower UTI.	[156]
	Finland	37 children with pyelonephritis due to *E. coli*. Patient blood samples 1–3 days following admission and 7 days following admission, and convalescence. ELISpot to detect IgM/A/G ASC to patient's own pathogen and P fimbria.	Mean IgM ASC 19 cells, IgA ASC 16 cells, IgG 6 cells /million PBMCs. IgM predominating in children <2 years of age. 33/37 (89%) children with detectable pathogen-specific ASCs. Responses correlated with increasing age, with four non-responders <1 year of age.	[157]

Influenza	UK (challenge study)	12 healthy adult volunteers. Blood samples at day -2, day 3, day 7 and day 28 relative to challenge. ELISpot to haemagglutinin to detect H1N1Br59-specific IgG ASCs.	Median Br59-specific IgG ASCs 166 (IQR 111–121)/million PBMCs at day 7. Antigen-specific ASCs were absent at other timepoints. Antigen-specific ASCs correlated with viral load and symptom score.	[158]
RSV	UK (challenge study)	61 healthy adult volunteers. Blood samples at day 0, day 3, day 7, day 10, day 14 and day 28 relative to challenge. ELISpot to detect RSV-lysate-specific IgM/A/G ASCs.	Infected volunteers showed median peak RSV-specific IgG ASCs of ∼300/million PBMCs at day 10. Antigen-specific ASCs were undetectable prior to day 7 and after day 14. RSV-specific IgG ASCs were largely undetectable in on-infected volunteers.	[159]
	USA	40 adult patients with RSV infection. Blood samples at enrolment, days 10–16 and days 22–45 following onset of symptoms. ELISpot to detect RSV-F protein-specific IgG ASCs.	RSV-specific IgG ASCs detectable by day 2 of illness, and detected in 90% of cases by day 11 (mean 200 RSV-specific IgG ASCs/million PBMCs). RSV-specific ASC responses were detectable in 48% of cases in convalescence (days 22–45).	[160]
	USA	97 adult patients with respiratory virus infection in total; data presented on 11 patients with RSV and 11 patients with influenza infection; and healthy controls including 19 adults recently vaccinated with against influenza, tetanus, HPV or HBV.	RSV-specific IgG ASCs detectable by day 2 of illness, and detected in 100% of cases by day 11 (median ∼1000 ASCs/million PBMCs) and absent in influenza infection and controls. Similarly, influenza-specific IgG ASCs detected in 100% of cases by days 4–11 (median peak ∼1000 ASCs/million PBMCs) and absent in RSV infection and controls.	[161]

		Patient blood samples at day 4–11 following symptom onset (enrolment), post-vaccine samples 6–7 days following vaccination. ELISpot to detect RSV-F protein-specific IgG ASCs, influenza antigens and vaccine antigens (as above).		
Dengue	Urban Thailand	46 adult patients with severe dengue infection. Patient blood samples on day of admission (median 6, range 2?8, days following onset of fever) and convalescence. ELISpot to detect dengue-virus specific IgM/A/G ASCs (using purified dengue virions).	Median peak dengue-specific IgG ASCs ∼15 000/million PBMCs (range 10–150 000/million PBMCs). Lower responses with IgA and IgM ASCs. Dengue-virus specific ASCs absent from controls and in convalescence.	[162]
	Brazil	84 child and adult patients with primary or secondary dengue infection, 15 controls with fever from other causes, and 10 local healthy controls. Patient blood samples on day of admission (median 6, range 1–9, days following onset of fever) and convalescence. ELISpot to detect dengue-virus specific IgG ASCs (using purified dengue virions).	Median peak dengue-specific IgG ASCs ∼20 000/million PBMCs (range 3000–100 000/million PBMCs), encompassing ∼50% of all IgG ASCs (to any antigen). Peak of IgG ASCs (to any antigen) between day 3.5 and day 7 following onset of symptoms, and correlated with onset of symptoms.	[163]
Mumps	USA (outbreak)	7 adult with acute mumps infection; 16 adult volunteers vaccinated with mumps virus vaccine (following previous infection/vaccination) Patient blood samples between day 7 and day 28 of onset of symptoms. Volunteer blood samples at day 0, day 7 and day 28 and 35 weeks ELISpot to detect IgG ASCs to purified mumps virions	Mumps virus-specific ASCs 10–100/million PBMCs from day 7–14 (6 patients), and 1.5/million PBMCs on day 28 (1 patient) Pre-vaccine and 35 weeks post-vaccine samples all mumps virus-specific ASCs <0.5/million PBMCs 100% sensitivity and specificity for diagnosis of acute mumps infection	[164]

Enterovirus-71 (EV-71)	Taiwan	28 children with acute EV-71 infection admitted to hospital Blood samples on admission (mean 4.7 days following onset of symptoms) and convalescence ELISpot to detect IgM/A/G ASCs to purified EV-71 virions	Mean EV-71-specific IgM ASCs 472, IgA ASCs 187, IgG ASCs 491/million PBMCs Responses largely undetectable from day 8 following onset of symptoms IgM ASC response predominated in younger children	[165]
Falciparum malaria	UK (challenge study)	9 adult volunteers initially vaccinated with experimental merozoite (blood-stage) vaccine, subsequently exposed to malaria challege infection Blood samples at day -1, day 7, day 11 and day 35 following malaria challenge infection ELISpot to detect MSP1-specific IgG ASCs	No detectable MSP1-specific IgG ASC responses at days 7, 11 and 35 in volunteers undergoing malaria challenge infection (despite merozoite parasitaemia by day 7) Possible questions regarding lack of later timepoints (since merozoite parasitaemia, not prehepatic sporozoites may act as immune stimulus)	[166]

Abbreviations: ALS, antibody from lymphocyte supernatant; ASC, antibody-secreting cells; CtxB, Cholera toxin B subunit; PBMCs, peripheral blood mononuclear cells; LPS, lipopolysaccharide; MSHA, mannose-sensitive haemogluttinin antigen; TB, tuberculosis; BCG, Bacille Calmette-Guérin; PPD, purified protein derivative; GMC, geometric concentration; ETEC, Enterotoxigenic *Escherichia coli*; CFA/II, *E. coli* colonising factor; CS1, CS3, CS6, coli surface antigens; LT, heat-labile *E. coli* toxin; UTI, urinary tract infection; OmpA, outer membrane protein A; RSV, respiratory syncytial virus; HPV, human papillomavirus; HBV, hepatitis B virus; EV-71, Enterovirus-71; MSP1, merozoite surface protein 1. Source: Carter et al. (2017)[121]. Used by kind permission of the authors.

ASCs in pneumococcal carriage and infection

Local mucosal and systemic adaptive immune responses occur secondary to NP carriage of pneumococci. Salivary IgA to pneumococcal proteins is secreted from adenoids isolated from healthy children with NP carriage of pneumococci[167]. In addition, increased serum concentrations of IgG to pneumococcal proteins are described in response to NP carriage of pneumococci in infants and young children[97–102]. There are currently no definitive data on whether serum antibody to pneumococcal antigens associated with NP carriage is entirely secreted by lymphoid tissue, long-lived plasma cells in the bone marrow, and/or whether pneumococcal antigen-specific ASCs are detectable in the blood of healthy children with NP carriage of pneumococci.

Pneumococcal pneumonia represents failure to contain pneumococci to the NP mucosa, leading to disease. A case series of 16 adult patients in Finland with bacteraemic pneumococcal pneumonia all had pneumococcal-specific IgG ASCs in peripheral blood on day 7 following detection of bacteraemia detected by ELISpot assay[153]. In all patients, pneumococcal-specific IgG ASCs could be detected, with IgM and IgA ASCs detected at lower frequencies. In 14 adult healthy controls IgG ASCs to the same pneumococcal antigens were all negative. In addition, total ASCs (i.e. not only pneumococcal-specific) were 10 times higher in cases than in controls, suggesting a significant polyclonal ASC response. Pneumococcal-specific ASCs derived from the same patients expressed a unique set of homing receptors in comparison to pneumococcal-specific ASCs following pneumococcal vaccination (n = 14 with pneumococcal polysaccharide vaccination, and n = 11 with pneumococcal conjugate vaccination). The authors suggest that pneumococcal-specific ASCs may be specific for pneumococcal infection and be derived from, and home to, the lung mucosa[168].

Work by the same group has recently characterised the antigen-specific ASC response to tonsillitis and sinusitis in 24 adults caused by *H. influenzae*, *S. pneumoniae* and Group A β-haemolytic streptococci[154]. Peripheral blood samples were taken 7–14 days following onset of symptoms. These patients all showed detectable IgG ASC responses specific to the causative pathogen, with slightly higher ASC frequencies in sinusitis than tonsillitis.

Taken as a whole, pneumococcal disease induces antigen-specific IgG ASC responses that are detectable in the peripheral blood of adults during illness. These ASC responses are present with or without detectable bacteraemia. *H. influenzae* and non-pneumococcal streptococcal respiratory mucosal infections induce a similar immune response. Therefore, ELISpot assay, may represent a useful new diagnostic test for the aetiology of pneumonia. Given that ALS has similar sensitivity and specificity to ELISpot assays for the detection of ASC responses in a number of infections (Table 1.3.2), ALS may also be useful for the diagnosis of pneumococcal pneumonia. No data are currently available formally assessing the utility of ALS for the diagnosis of pneumococcal infections.

1.4 The transcriptomic response to infection

Fundamentally the immune response to infection is a result of changes in the expression of genes, and the proteins that they encode. Transcriptomics has been applied to three problems within the study of the immune response to infection. Firstly, transcriptomic analysis may reveal pathogen or pathogen class-specific signatures that may diagnose the aetiology of infection[169]. Analysis of the whole transcriptome is currently too cumbersome for routine clinical care, but the identification of clinically-relevant transcriptomic signatures based on a small number of genes may have revolutionary potential for laboratory or point-of-care diagnostics. Secondly, the transcriptome provides a comprehensive view of the cellular processes at a particular point in time, allowing bioinformatic analysis of pathways that may be activated by individual pathogens and in differing disease states[170]. This is particularly powerful when combined with knowledge of gene function to provide a comprehensive understanding of the pathophysiology of infection and immune responses in each patient[171, 172]. Finally, and not discussed further here, the study of the BCR repertoire at both an individual and population level using RNA sequencing may define the molecular evolution of the adaptive immune response in absolute detail over the coming decades[173–175].

1.4.1 Transcriptomic techniques

RNA expression microarrays developed rapidly in the early 2000s, following completion of the initial builds of the human genome. The initial microarrays were created by the binding of either cDNA complementary to a gene, or oligonucleotides complementary to that gene, to a glass or polymer slide in an exact (and known position). When multiplied over the thousands of annotated genes in the genome (in this case, human genome) the microarray represents a grid, with each locus a gene. RNA from a sample is fluorescently labelled and then incubated with the array, with binding of complementary sequences, before any unbound RNA is washed away[176]. The relative concentration of RNA specific to a gene is then interpreted from the fluorescence of the corresponding locus on the array.

Until recently, microarrays were the standard manner in which to assess the human transcriptome. Indeed, almost all studies discussed in this section are based on microarrays. Microarrays require a number of quality controls steps to remove low-quality data for probes and samples, followed by normalisation methods to adjust for experimental variability, such as batch effects from different chips. Microarrays are considerably cheaper than the alternative, sequencing of RNA, and can be designed for a wide variety of experiments. Nonetheless, there are considerable limitations to microarrays, including difficulty quantifying low level gene expression (due to background fluorescence), the limited dynamic range of arrays (thereby underestimating log-fold changes in gene expression), and the inability of arrays to uncover novel features (since only "known" genes are included on the microarrays).

Given these limitations, there has been a recent move towards investigating the transcriptome by direct sequencing of RNA[177]. A variety of methods may be used which primarily differ in the accuracy of "reads" for the transcript (for example, the number of erroneous bases), and the length of these reads. A typical method (Illumina HiSeq4000, used to generate the data in this thesis) sequences the transcriptome in "reads" of 75 base pair lengths. These reads may then be aligned to their corresponding position on the human genome by matching of the nucleotide sequences – a bioinformatic challenge given

that the transcriptome consists of large numbers of splice junctions[178]. Quantitation of gene expression follows, by the "counting" of reads that are aligned to genes in the genome[179]. There is enormous breadth to the possibilities that RNA sequencing offers including the building of *de novo* transcriptomes, in addition to a wide dynamic range and lack of background hybridisation. Despite these potential advantages, recent experiments to assess RNA sequencing vs microarray technology for clinical endpoint prediction have showed that there is little difference in the overall results[177].

Bioinformatic methods are progressing rapidly to meet the demands of complex analyses specific to sequencing data[180]. However, assessment of differential gene expression, where the data are described as "counts" of genes, is similar regardless of whether the data was created by microarray experiments or through RNA sequencing[181]. Despite this, there are likely to be significant bioinformatic challenges in transposing the results of microarray-based studies to new data that is likely to increasingly be derived from RNA sequencing.

1.4.2 Transcriptomics for the diagnosis of infectious diseases

Analysis of host genomewide gene expression may reveal differential gene regulation between groups with differing aetiologies of infection[182]. This has termed the transcriptomic (or RNA) signature of an infection. As with studies of novel diagnostic tests (detailed above) a typical study design would assess the ability of a test (i.e. the differential regulation of a set of genes) to discriminate between "definite cases" and "definite controls" (typically children unwell with another infection), and then to apply an optimal test threshold to the intermediate group. The use of a high-throughput whole genome approach is relatively unbiased in its selection of transcripts to distinguish between aetiologies; it may also reveal previously unrecognised immunological pathways. The enormous number of statistical tests that are undertaken in analysis of differential gene regulation has brought new statistical challenges, necessitating the development of a "false discovery rate" (FDR) for the interpretation of a "significant" result.

Recent work has shown that differential regulation of a very small set of genes can accurately diagnose bacterial infections in children and neonates[112, 183–185] with greater accuracy than other markers of infection such as procalcitonin[186–188]. A two-gene transcriptomic signature to differentiate bacterial from viral infection in a discovery group of 240 children, and a test group of 130 children (median age of 19 months) with a large range of bacterial, viral and indeterminate infections has been described[112]. In this signature, upregulation of the gene family with sequence similarity 89 (*FAM89A*) was associated with bacterial infection. This gene was inversely regulated in viral infections, with interferon-induced protein 44-like (*IFI44L*) upregulated. A test based on the relationship between *FAM89A* and *IFI44L* gave a sensitivity of 100% and a specificity of 95% for the diagnosis of bacterial infection in the test set, a sensitivity of 100% and specificity of 86% in a separate validation set. Without further optimisation, this test was also able to discriminate between bacterial and viral infections in neonates with a sensitivity of 89% and specificity of 94%. With the rapid advance of transcriptomic technology (and reduction in cost) this raises the prospect of transcriptomic results being available in a timescale that is clinically useful across a range of settings to diagnose bacterial infection in children with an infectious disease syndrome.

One might hypothesise that differentiating between closely related pathogens on the basis of the host transcriptome provides a greater challenge than differentiating between bacterial and viral infections[189]. As an example, transcriptomic methods have been used to discriminate between latent tuberculosis and and active tuberculosis in adults[190] and between other infections, latent tuberculosis and active tuberculosis in African children[191]. In African children, a 51-gene transcriptomic signature was able to discriminate between tuberculosis and other infections with a sensitivity of 83% and specificity of 84% for the diagnosis of culture-confirmed tuberculosis. Other pathogens for which a specific transcriptomic signature has been proposed include sepsis due to *Burkholderia pseudomallei* infection versus other sepsis aetiologies in adults[192], sepsis with Gram-negative vs Gram-positive bacteria in children[193], influenza from healthy controls[194], sepsis with *Staphylococcus aureus* vs *E. coli* in adults[195], and sepsis with bacterial infection in neonates vs non-infected neonatal controls[196, 197]. In addition,

data from few (typically <60) gene transcripts are required to accurately discriminate between influenza and other respiratory infections[198, 199]. The severity of infection and prognosis of children with severe viral infections[200], adults with pneumonia[201], and with sepsis[171, 172] can also be inferred accurately.

Despite this enormous progress, the small number of patients with microbiologically-confirmed infection with a specific pathogen, varying criteria for the enrolment of control groups, and differing technical and bioinformatic methods hamper the integration of pathogen-specific transcriptomic signatures across datasets (with some exceptions[184, 188]). In addition, the enormous biological variability inherent within adults, children and neonates – even when infected with the same pathogen – may limit the accuracy of these signatures when prospectively applied to unselected case series of febrile patients. Such variability may result from transcriptional instability during the course of illness[202] or from genetic or epigenetic mechanisms in the host response to infection[203].

1.5 Aim and objectives

Within the broad context described above, I investigated childhood pneumococcal pneumonia in Nepal. I initially collated and examined data from a long-term surveillance programme of childhood invasive bacterial disease at Patan Hospital, Kathmandu, Nepal, with a particular focus on invasive pneumococcal disease. I then examined childhood pneumonia in the context of detailed case series at Patan Hospital, focusing on describing the burden of pneumococcal pneumonia in the context of introduction of 10-valent PCV. I then explored two novel diagnostic tests for the aetiology of childhood pneumonia: assay of ALS, and diagnosing pneumonia aetiology based on the transcriptomic response to infection.

The work detailed within this thesis had four aims:

1. To describe the burden of pneumococcal infection in children with suspected invasive bacterial disease and pneumonia admitted to Patan Hospital, Kathmandu Valley, Nepal (Chapter 3).

I aimed to describe the prevalence of invasive bacterial disease, and particularly invasive pneumococcal disease, in children admitted to Patan Hospital with suspected pneumonia, meningitis or sepsis. I used data arising from a decade-long programme of surveillance for childhood invasive bacterial disease at Patan Hospital (March 2005–December 2016). In addition, I aimed to detail the clinical epidemiology of pneumonia in children at Patan Hospital, with a particular focus on diagnosing pneumococcal pneumonia, in a case series of children admitted to Patan Hospital with a diagnosis of pneumonia between March 2014 and December 2016 (the Inpatient Pneumonia Study).

2. To assess the impact of the introduction of 10-valent PCV on the clinical epidemiology of childhood pneumonia at Patan Hospital (Chapter 4).

I aimed to describe and analyse changes in the clinical epidemiology of childhood pneumonia at Patan Hospital that were associated with the introduction of 10-valent PCV to the infant immunisation schedule in Nepal (reaching the Kathmandu Valley by August 2015). I used data from the Inpatient Pneumonia Study.

3. To investigate the utility of assay of ALS to pneumococcal antigens for the aetiological diagnosis of pneumococcal pneumonia in children (Chapter 5).

I developed and managed a study, termed the Pneumonia Aetiology in Nepal: Diagnosis with Immunology and Transcriptomics (PANDIT) study. This had considerable overlap of patients with the Inpatient Pneumonia study, but was designed to assess novel diagnostic tests for the aetiology of childhood pneumonia. Assay of ALS to pneumococcal antigens (both capsular polysaccharides and proteins) was investigated.

4. To investigate the transcriptomic response to pneumonia in children in Nepal to assess the utility of transcriptomic analyses to discriminate between pneumococcal pneumonia and other aetiologies (Chapter 6).

To investigate the transcriptomic response to pneumonia, I used samples and data arising from the PANDIT study. I used analysis of differential gene expression to develop a method to discriminate between pneumococcal pneumonia and other aetiologies.

The methods used for these investigations are detailed in Chapter 2 and in the Methods section of Chapters 3, 4, 5, 6. Chapter 7 summarises the results of each of these analyses.

2

General methods

2.1 Study site and population

Geography and human development

The Federal Democratic Republic of Nepal is a small landlocked South Asia country. Its southern, eastern and western borders are with India, and its northern border – via the Himalayas – is with Tibet, an autonomous region of the People's Republic of China. The population of Nepal was estimated to be 26.5 million in 2011 (the most recent census), with 10% of people <5 years of age. Approximately 4.5 million people (17%) lived in urban areas, mainly in the Kathmandu Valley, and this number is increasing rapidly[204].

The socio-economic status of Nepal was ranked 144 of 188 countries – medium human development – by the UN Development Programme in 2015 using the Human Development Index, a composite of life expectancy at birth, years of schooling and gross national income per capita[205]. Life expectancy at birth was 70 years of age in

2015. However, there is considerable variability within Nepal in human development between men and women, ethnic groups/castes, and geographical regions, particularly with regard to access to education – although these inequalities appear to be reducing over time[206].

Child health

The mortality rate of children <5 years of age in Nepal is estimated to be 37 per 1000 births, of which 22 per 1000 are still births or neonatal deaths[7]. This represents a 5.2% annual decrease between 2000 and 2015, and compares to a mean mortality rate of children <5 years of age of 50 per 1000 births in the South Asian countries of Bangladesh, Bhutan, India, Nepal and Pakistan. Until 2015, the Nepali infant immunisation schedule included vaccination with antigens of diphtheria-tetanus-pertussis, BCG, hepatitis B, Hib (from 2009), oral polio virus (with the addition of inactivated poliovirus in 2014) and measles. The most recent Demographic Health Survey (2011) estimated that 87% of children <24 months of age had received all of these vaccinations, ranging from 85% in the lowest wealth quintile to 96% in the highest wealth quintile, and with little variation in coverage between rural (87%) and urban (90%) regions[207].

Other indicators of child health are diverse within Nepal. Approximately 40% of children <5 years of age were malnourished (weight-for-age Z-score <2 standard deviations below the WHO-defined mean) in the lowest wealth quintile ranging to 17% in the highest wealth quintile, with significant variation between rural (30%) and urban regions (17%). Across all indicators (including vaccination, and antenatal/perinatal care) lack of formal education for mothers is most clearly associated with poor health indicator status[207]. Prevalence of HIV infection across Nepal is estimated at <1%, although this is also estimated to be considerably higher in marginalised and high-risk populations, such as street children in the Kathmandu valley, sex workers and men who have sex with men[208].

Patan Hospital, Kathmandu, Nepal

The Kathmandu Valley is a generally urbanised area in the middle hills of Nepal, consisting of Kathmandu, Lalitpur (including Patan) and Bhaktapur. Human Development Indices are significantly higher in the Kathmandu Valley than any other district in Nepal.

Patan is the major city of the Lalitpur sub-Metropolitan District of Nepal, which is contiguous with the city of Kathmandu immediately to the north. Patan Hospital is one of the largest hospitals in Nepal, and one of a small number with inpatient paediatric facilities, including a paediatric intensive care unit, in the Kathmandu Valley. Patan Hospital is a not-for-profit organisation: modest fees are paid by most patients for healthcare, although these are waived for patients if necessary. It is also the base hospital for Patan Academy of Health Sciences – a medical school and academic medical centre (http://www.pahs.edu.np).

A large number of private pharmacies operate in Patan with ready-availability of medications – including a wide range of antibiotics – without a prescription.

The Nepal earthquakes of 2015 and political unrest

Nepal, and specifically the Kathmandu Valley and areas immediately to the north and east of the Kathmandu Valley suffered a series of large earthquakes in April and May 2015. As an immediate result of the earthquake more than 8000 people were killed, 22 000 injured, and approximately 2.5 million people made homeless [209]. The Kathmandu Valley witnessed large fluctuations of people during this period with families travelling from the Kathmandu to rural areas for rebuilding, and internally-displaced people's camps arising in Kathmandu for the homeless ([210] and *personal observations*).

In addition, political unrest and difficulties importing petroleum from India between September 2015 to February 2016 led to the use of alternative fuels (low quality diesel, wood, cow dung) in Kathmandu and elsewhere. In addition, dust from demolition and construction work in the Kathmandu Valley may have increased the levels of dust particulates in the air. This resulted in a large increase in air pollution in the Kathmandu Valley[211]. Displacement of large numbers of people from their homes, and increased air pollution may have increased the incidence of pneumonia by facilitating the transmission of respiratory pathogens, as has been seen in other crisis-affected populations[212, 213].

2.1.1 Introduction of PCV to the Nepali infant immunisation schedule

10-valent PCV was introduced to the infant immunisation schedule of Nepal during 2015, and was implemented in the Kathmandu Valley on 21st August 2015. The routine schedule gives 10-valent PCV at 6 weeks and 10 weeks of age, followed by a booster dose at 9 months of age. There was a limited "catch-up" campaign of one or two doses of 10-valent PCV for children >10 weeks of age but <1 year of age until the birth cohort reached nine months of age.

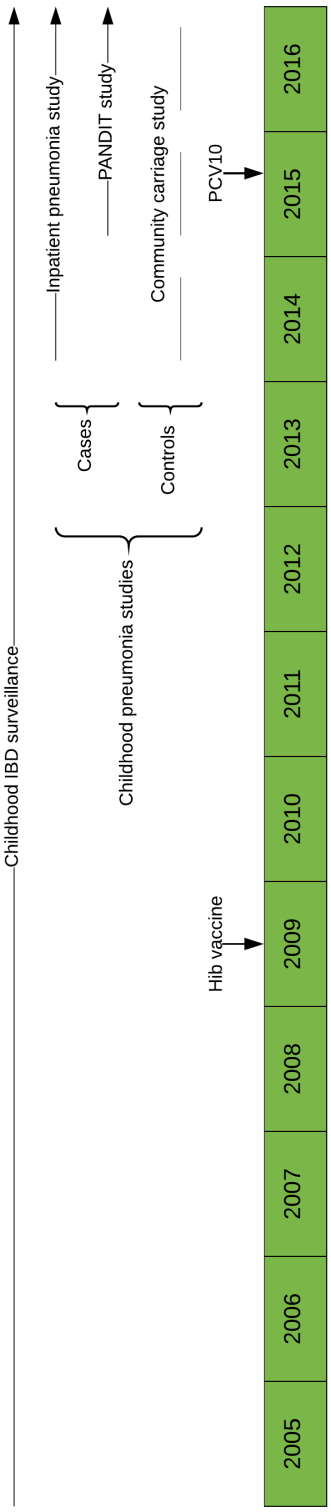
2.1.2 Overview of case definitions used within these studies

This thesis describes data from four linked studies: childhood invasive bacterial disease (IBD) surveillance (including admitted children with suspected pneumonia, meningitis or sepsis), the Inpatient Pneumonia Study and the Pneumonia Aetiology in Nepal: Diagnosis with Immunology and Transcriptomics (PANDIT) study (of cases of children with pneumonia) and the Community Carriage Study (of children in the community without pneumonia or other suspected serious febrile illness). The time course for enrolment into these studies is described in Figure 2.1, with inclusion criteria and overlap of patients described in the text.

2.2 Childhood invasive bacterial disease surveillance

2.2.1 Definitions, inclusion criteria and clinical methods

Since March 2005, all children aged 0 months to 14 years admitted to the Paediatric Ward at Patan Hospital with suspected community-acquired IBD have been surveyed. Suspected IBD encompassed all children with clinically suspected pneumonia, clinically suspected meningitis, or clinically suspected sepsis (Table 1.1) as determined by the admitting paediatric team, and based on guidelines within the WHO *Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illness*[36].



Childhood invasive bacterial disease surveillance: Children aged 0 to 14 years with suspected pneumonia, meningitis or sepsis at presentation and admission to Patan Hospital. Consent was not required for this surveillance. All children within the inpatient pneumonia study and the PANDIT study were therefore automatically enrolled to childhood IBD surveillance.

Inpatient pneumonia study: Children aged 0 to 14 years with suspected pneumonia at presentation and admission to Patan Hospital. Consent was required from parents.

PANDIT study: Children aged 0 to 14 years with suspected pneumonia at presentation and admission to Patan Hospital, and from September 2015 to December 2016 a small number of children who were treated for pneumonia at Patan Hospital but not subsequently admitted. Consent was required from parents. All admitted children in the PANDIT study were also enrolled in the inpatient pneumonia study; non-admitted children in the PANDIT study were not able to be enrolled in the inpatient pneumonia study.

Community carriage study: Children from the community attending Patan Hospital for routine reasons, or siblings of children admitted to Patan Hospital (for any reason), who did not have a diagnosis of pneumonia. Children with mild upper respiratory tract infection were eligible for enrolment. Enrolment was age-stratified into children 6 months to <2 years of age and 2 years to <5 years of age. Consent was required from parents.

Figure 2.1: Overview of time course for enrolment, and inclusion criteria, for studies described in this thesis.

From March 2005 to end of December 2006, detailed clinical data from all children with suspected IBD was recorded. From 2007 to July 2009 only data from cases of children with bacteria isolated from normally sterile sites (i.e. blood, pleural fluid or CSF) were recorded. Since August 2009 all new admissions to the Paediatric Ward (of any presentation) have been recorded daily, with a detailed case record form completed (including demographic, clinical and laboratory data) on all children admitted with suspected IBD. These data form the basis of the childhood surveillance programme at Patan Hospital.

2.2.2 Microbiology: cultures from normally sterile sites

Blood was taken aseptically from all children with suspected IBD for culture on admission to Patan Hospital, inoculated into a Bactec Peds Plus culture bottle (Becton Dickinson, USA) and incubated aerobically in 5% CO_2 at 35–37°C in the Microbiology Laboratory at Patan Hospital. From 2005–2009 incubation required venting of blood culture bottles and the use of a standard incubator. From 2009 onwards an automated blood culture system was used (Bactec, Becton Dickinson). Subculture of turbid samples was onto sheep blood (obtained weekly from sheep at Patan Hospital) and chocolate agar in the first instance. Organisms cultured from blood were identified using standard microbiological and biochemical methods[214], and antimicrobial sensitivities were tested using the disc diffusion method[215]. Six-monthly audits of blood culture volumes were undertaken and presented in order to support good practice in blood culture techniques.

Samples of body fluid from other normally sterile sites (in practice, pleural fluid or CSF) were immediately plated onto Columbia agar plates containing 5% sheep blood and incubated as above.

2.3 Childhood pneumonia studies

In addition to detailing data on childhood IBD surveillance at Patan Hospital (Section 2.2 and Chapter 3), this thesis presents data from three inter-linked studies:

1) The Inpatient Pneumonia Study

2) Pneumonia Aetiology in Nepal: Diagnosis with Immunology and Transcriptomics (the PANDIT study)

3) The Community Carriage Study

The inclusion criteria, exclusion criteria and and recruitment process were similar for both the Inpatient Pneumonia Study, and the PANDIT study. However, the Inpatient Pneumonia Study commenced enrolment in March 2014, and the PANDIT study commenced enrolment in February 2015. In addition, whilst the Inpatient Pneumonia Study was strictly limited to children admitted to Patan Hospital, the PANDIT study enrolled a small number of children who were diagnosed with pneumonia at Patan Hospital, but not admitted to the hospital. Methods for these two studies are described below.

The Community Carriage Study was a study of the NP carriage of pneumococci in children in the community of Patan. As such, the inclusion criteria, exclusion criteria and recruitment process were separate from the two other studies. They are detailed below, within this section. In this thesis I present only a small amount of data from the Community Carriage Study, since this is necessary to provide community controls for the serotype-specific pneumococcal NP carriage of children in the Inpatient Pneumonia Study.

2.3.1 Definitions, inclusion criteria and clinical methods

Children $\geq$2 months (60 days) of age and up to 14 years of age who were being admitted to Patan Hospital with a clinical diagnosis of pneumonia were eligible for the Inpatient Pneumonia Study and the PANDIT study. A child was deemed to have a clinical diagnosis of pneumonia if the admitting paediatric team (including consultant paediatrician) gave a diagnosis of pneumonia. The diagnosis was based on clinical information, and was typically made before results of chest radiographs and blood neutrophil count were available to the clinical team. Plasma CRP concentrations were not available at Patan

Hospital. Recruitment to the studies was following informed, written, consent following discussion with a Clinical Research Fellow at Patan Hospital. The Clinical Research Fellow had no role in the clinical diagnostic process for each child. A Clinical Research Fellow was available to recruit patients between 08:00 hours and 20:00 hours on every day of the studies.

The Inpatient Pneumonia Study

The Inpatient Pneumonia Study (detailed in Chapters 3 and 4) commenced enrolment in March 2014, and is currently continuing. Data for the Inpatient Pneumonia Study are shown until the end of December 2016. Only children who were admitted to Patan Hospital were eligible for the Inpatient Pneumonia Study. All children enrolled into the Inpatient Pneumonia Study were included in the childhood IBD surveillance programme at Patan Hospital, and had blood cultures taken as described in Section 2.2.2. In addition, all children enrolled into the Inpatient Pneumonia Study had culture of a NP swab to detect serotype-specific pneumococcal carriage (see Section 2.3.2) and a chest radiograph at admission/recruitment. Specific clinical methods for the Inpatient Pneumonia Study are described in Chapter 3.

The PANDIT study

The PANDIT study commenced enrolment on 26th February 2015. Within the PANDIT study, children were enrolled for investigation of antibodies from lymphocyte supernatant (ALS) for the detection of pneumococcal responses from antibody-secreting cells (ASCs) until 30th September 2016 (see Chapter 5); data for this section of the PANDIT study are included here. Within the PANDIT study, enrolment of children for investigation of the transcriptomic response to childhood pneumonia has continued since February 2015; data for this section of the PANDIT study are included until end of December 2016 (see Chapter 6).

The large majority of children who were enrolled into the PANDIT study were also enrolled to the Inpatient Pneumonia Study. However, since September 2015, a small number children with a clinical diagnosis of pneumonia made in the Outpatient Department of Patan Hospital, and not admitted to Patan Hospital, have been enrolled (see Chapters

5 and 6 for details). All children enrolled into the PANDIT study (with the exception of healthy controls) had blood cultures taken as described in Section 2.2.2, and culture of a NP swab to detect serotype-specific pneumococcal carriage. In addition a number of healthy controls were enrolled into the PANDIT study to provide an additional comparator group for the analysis of assays of ALS for the detection of pneumococcal responses from ASCs, and transcriptomic investigation of childhood pneumonia (see Chapters 5 and 6 for details). Only children enrolled to the PANDIT study had plasma CRP concentration measured (retrospectively, excepting healthy controls), and blood samples taken for assay of ALS and transcriptomic analysis. Specific clinical methods for the PANDIT study are described in Chapter 5.

2.3.2 Microbiology: cultures of nasopharyngeal swabs and serotyping

A single nylon flocked NP swab (Thermo Fisher Scientific, UK) was obtained from each case in the Inpatient Pneumonia Study or PANDIT study by a trained Clinical Research Fellow and aseptically placed directly into a sterile tube containing 1 ml of skim-milk-tryptone-glucose-glycerin (STGG) medium using WHO standardised methodology[216]. Swab specimens were immediately taken to the Microbiology Laboratory at Patan Hospital and cultured as follows. The tube containing the NP swab and medium was briefly vortexed before a 200 μl aliquot was taken aseptically and plated onto Columbia agar plates containing 5% sheep blood and incubated overnight at 37°C. The remaining medium and swab were stored in the original tube at -80°C at Patan Hospital. Any colonies of *S. pneumoniae* were identified from the agar plates on the basis of morphology and optochin susceptibility; bile solubility was not assessed on bacterial isolates from the nasopharynx.

A single colony of pneumococci was taken for serotyping (or two colonies if the morphology differed). Serotyping was done via the Quellung method using specific-sera to pneumococcal polysaccharides (Statens Serum Institut, Denmark) according to agreed guidelines[41]. All colonies were serotyped at Patan Hospital by laboratory technicians (or a Clinical Research Fellow, MJC) trained and experienced in the Quellung method.

Table 2.1: Standarised interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies[217]

Definition	Description
Normal	No consolidation/effusion/infiltrate
Other infiltrate	Linear and patchy densities (interstitial infiltrate)
Endpoint consolidation	Dense or fluffy opacity occupying at least a portion of a lung, or pleural effusion (with or without infiltrate)

Pneumococcal isolates of presumed rare serotypes, for which specific sera were not available at Patan Hospital, were serotyped at Oxford Vaccine Group following transport of samples on dry ice. In addition, a quality control check was provided by serotyping of a random 10% of other isolates at Oxford Vaccine Group.

2.3.3 Radiology: interpretation of chest radiographs

All children recruited to the Inpatient Pneumonia Study and the PANDIT study (excepting children enrolled as outpatients, and healthy controls) had a chest radiograph done on admission/recruitment to the study. Digital radiographs were done in the Department of Radiology, Patan Hospital, and recorded on a server specific to the Department. At approximately 3-monthly intervals each radiograph was read independently, and blinded to patient details, by a paediatrician (MJC) and a radiologist. Both readers were trained in the standardised interpretation of chest radiographs for the diagnosis of pneumonia in children in epidemiological settings[30]. Readings which differed, and a random 10% of all chest radiographs were referred to a senior radiologist for arbitration. The same two readers, and single arbiter, were used throughout the study. Inter-observer agreement was audited approximately every 200 radiographs, feedback provided, and further training provided as necessary by the arbiter. The criteria for the interpretation of chest radiographs are described in Table 2.1, with typical features shown in Figure 2.2.

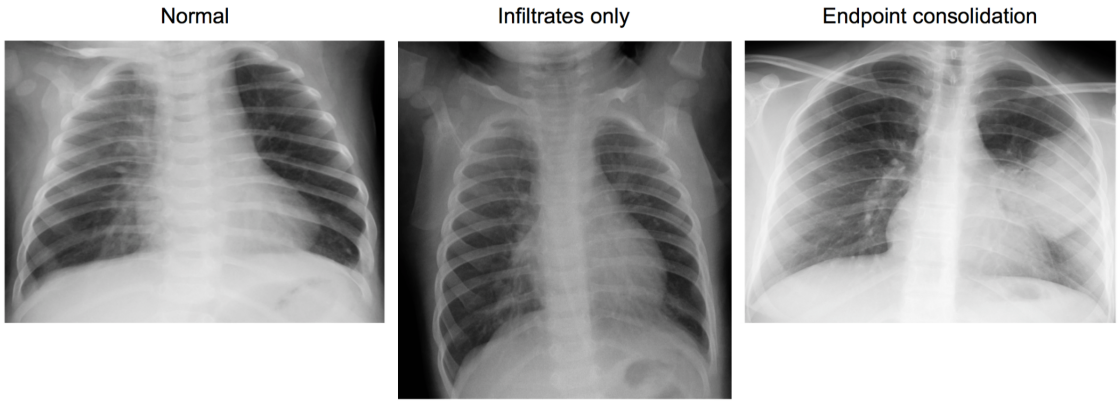


Figure 2.2: Examples of a "normal" radiograph, a radiograph with "other infiltrate only", and a radiograph with "endpoint consolidation" from children enrolled in the Inpatient Pneumonia Study at Patan Hospital.

2.3.4 The Community Carriage Study

The Community Carriage Study aimed to assess the prevalence of serotype-specific pneumococcal NP carriage in children in the community of Patan prior to, and following, the introduction of 10-valent PCV.

Children were enrolled in an age-stratified manner in three groups: 0 to <8 weeks of age, 6 months to <2 years of age, and 2 years to <5 years of age. Data from children 0 to <8 weeks of age are not included in this thesis. Children were eligible for enrolment if they were of the appropriate age, and did not have a diagnosis of pneumonia, and signed and witnessed parental consent was given. The Nepal Health Research Council and Oxford Tropical Research Ethics Committee agreed the study protocol.

Of children 6 months to <2 years of age, 1149 children were enrolled between April and September 2014, 595 children were enrolled between April and September 2015 and 1151 children were enrolled between April and September 2016. Of children 2 years to <5 years of age, 156 children were enrolled between April and June 2014 and 151 children were enrolled between April and June 2016.

All children enrolled to the Community Carriage Study had detailed demographic data, and vaccination status (for all routine vaccinations), recorded. A single nylon flocked NP swab was obtained from each case in the Community Carriage Study and cultured

for pneumococci as described above (Section 2.3.2). Serotyping was by the Quellung method, using identical methods (including quality control) as described above.

2.3.5 Data management and analysis

Data were managed on study-specific databases in Microsoft Access. Data cleaning and analysis was undertaken on Stata 14 (Statacorp, USA) and the R statistical environment[218]. Details of analyses are given in the relevant chapter.

2.4 Methods specific to the PANDIT study

Approximately 3 ml of venous whole blood was sampled from each child at recruitment to the PANDIT study. Of the 3 ml of blood sampled at each time point, 2 ml was transferred immediately into a heparinised 15 ml centrifuge tube and processed immediately to generate ALS (see below, Section 2.4.1). Remnant plasma from the processing of this aliquot of blood was labelled and immediately stored at -80°C, prior to transport on dry ice to the Biochemistry Laboratory at Oxford University Hospitals NHS Foundation Trust for measurement of CRP concentrations.

The remaining 1 ml of blood was transferred immediately to an RNA-stabilising tube, labelled, and stored at -80°C at Patan Hospital prior to shipment on dry ice to the Wellcome Trust Centre for Human Genetics (WTCHG), University of Oxford, for the extraction and processing of RNA (see below, section 2.5.1).

As noted above (Section 2.3.1), blood from children enrolled to the PANDIT study between February 2015 and September 2016 was processed to generate ALS and stored for further analyses of RNA. Blood from children enrolled from October 2016 onwards was used for measurement of plasma CRP concentrations and further analyses of RNA only.

At discharge from Patan Hospital, all children participating in the PANDIT study were invited back to the hospital for a convalescent sample 6–8 weeks following their initial

sample (convalescent time point). Approximately 45% children returned for convalescent sampling.

2.4.1 Methods to generate ALS

2 ml of fresh whole blood from children enrolled into the PANDIT study was separated into plasma, PBMCs and a mixture of red cells and polymorphonuclear cells within 4 hours of sampling as follows.

5 ml of sterile-filtered Ficoll and sodium diatrizoate, with a density of 1.077 g/ml (Histopaque 1077, Sigma-Aldrich, USA), was added to a sterile 15 ml centrifuge tube. Fresh whole blood was carefully layered onto the Ficoll using a sterile Pasteur pipette and the tube was centrifuged at 400 g for 20 min with low acceleration and the brake off. Following this, plasma was removed (the upper layer), labelled and stored at -80°C. The PBMC layer was then removed using a new sterile pipette, and placed into a new sterile 15 ml centrifuge tube, with the additional of 5 ml of R0 culture medium (see below). The tube was centrifuged at 2000 RPM for 10 mins, before the R0 was carefully removed from the PBMC pellet (wash 1). The cell pellet was carefully resuspended in a fresh 5 ml volume of R0, with the removal of a 50 μl aliquot of PBMC suspension for counting, and a second centrifuge step (wash 2). The number of PBMCs in 50 μl of cell suspension was calculated by manual counting of PBMCs stained with Trypan Blue 0.4% (VWR International) using a haemocytometer. Following wash 2, R0 was carefully removed from the PBMC pellet, and the pellet resuspended in R10 medium to a density of 1 x 10^7 PBMCs/ml.

The resulting cell suspension was transferred to a single well of a 24 well cell culture plate (Greiner Bio-one, Germany), with phosphate buffer saline (PBS) added to adjacent cells to minimise evaporation, covered with the lid and incubated at 5% CO_2 and 35–37°C for 48 hours.

Following incubation, ALS was harvested by transfer of the cell suspension to a micro-centrifuge tube and centrifuged at 13 000 RPM for 10 min. The resulting supernatant (ALS) was separated from the PBMC pellet and transferred to a sterile cryovial. A small

volume of 40X concentration of protease inhibitors (approximately 25 μl per 1 ml of ALS) was added to the ALS to prevent proteolysis, and the ALS was labelled and stored at -80°C at Patan Hospital.

All steps in this method were done in a strictly aseptic manner, with a PCR cabinet used for transfer of materials between tubes. Media were prepared in a grade 2 Biohazard Safety Cabinet at Patan Hospital. The media used in these methods are detailed in Table 2.2.

The full protocol for the generation of ALS is included in the Appendix, Figure A.1.

2.4.2 Antibodies to pneumococcal polysaccharides by fluorescent multiplexed bead-based immuno-assay

See Chapter 5 for methods.

2.4.3 Antibodies to pneumococcal proteins by fluorescent multiplexed bead-based immuno-assay

See Chapter 5 for methods.

2.5 Transcriptomic investigation of childhood pneumonia: the PANDIT study

2.5.1 Sampling, extraction and quality control of RNA

Fresh whole blood samples for analysis of RNA were immediately inoculated into RNA stabilising reagents in a purpose-designed collection tube (Tempus Blood RNA Tube, Applied Biosystems, USA) and and stored at -80°C within 4 hours of sampling. Samples for analysis of RNA were shipped to the WTCHG on dry ice in two separate batches in June 2016 and January 2017.

RNA was extracted from Tempus Blood RNA Tubes using the Tempus Spin Isolation Kit, according to the manufacturer's protocol[219]. Optional modifications to the protocol were: the addition of an extra volume of PBS to ensure a consistent volume of blood/reagent volume of 12 ml (step 3 of the protocol) prior to vortexing, and the use of a DNase treatment (step 9 of the protocol) using AbsoluteRNA Wash Solution (Applied Biosystems). The final volume of RNA eluate was 90 μl. This volume was sent to the High-Throughput Genomics Group at the WTCHG. Here, total RNA was normalised to 750 ng and libraries prepared using Illumina TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Globin (RS-122-2503, High Throughput, Illumina, USA). Libraries were pooled and sequenced on HiSeq4000 at 75 bp paired end. All samples passed basic quality control steps (e.g. for the depletion of ribosomal RNA), and were sequenced to a depth of at least 28 million reads. The data was made available as fastq files which were downloaded and analysed on the servers of the WTCHG.

I thank the High-Throughput Genomics Group at the WTCHG (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of sequencing data, and the Research Computing Team at the WTCHG for provision of computing facilities.

2.5.2 Bioinformatic methods for RNA-sequencing data

Following RNA-sequencing, the raw data for the 75 bp paired end reads were uploaded to the computing clusters at the WTCHG, where all subsequent bioinformatic analyses took place. Raw reads were aligned to the human genome (build GRCh38, available from http://www.gencodegenes.org/releases/21.html) using the alignment tool Bowtie2 (as part of the Tophat package)[178]. This software is widely used for comparative genomics and transcriptomics, and is specifically designed to tolerate gapped alignment (as is necessary for mapping only transcribed genes). The script, including parameters used for Bowtie2, are available on the J Knight Group github pages (see Appendix). Following alignment, reads were filtered to remove duplicated reads using SAMTools[220]. The script is available as above.

Following alignment and filtering, I used the software HTSeq[179] to generate counts of the number of filtered reads mapped to each gene isoform within the human genome build. These data were imported into R[218] for further analyses using the package $edgeR$[181, 221]. The numbers of reads mapped to each gene were normalised to the length of the transcript for each gene isoform, and the total number of reads for each sample using trimmed means of M-values[222]. For these analyses, only protein coding genes from autosomes were used (i.e. pseudogenes and other genomic features were removed) and gene isoforms were collapsed into their parent genes. Genes that were expressed in fewer than 8 samples (i.e. the number of children with definite pneumococcal pneumonia) were removed from analyses.

Differential gene expression between individuals, and within individuals but between acute and convalescent time points, was calculated as implemented in $edgeR$, which assumes a negative binomial distribution of gene counts and uses a test analogous to the Fisher Exact Test, adapted for overdispersion of data[181]. I initially used principal component analyses to inspect the first, second, third and fourth principal components of the variance in gene expression amongst samples. I used the R package $ellipse$ to plot the 95% confidence interval around a cluster of samples[223]. This package uses a bivariate (e.g. principal components 1 and 2) normal distribution about the calculated centre of a defined (supervised) cluster[224]. Its reliance on the normal distribution makes it highly conservative for datasets with small numbers of samples in each cluster (as here).

Gene pathway analysis was done by analysing the list of differentially expressed genes for each comparison in the tool *eXploring Genomic Relations*, available as the R package XGR[225]. This tool uses prior biological knowledge (here, derived from the KEGG/REACTOME/BioCarta pathway ontologies, *http://www.genome.jp/kegg/*)to analyse the networks associated with each differentially expressed gene in a tree-like structure. Enriched (upregulated or downregulated) pathways from the "root" to the terminal "branches" of the tree are determined. I set the maximum number of genes to be annotated to 2000, and compared differentially expressed genes arising from comparisons made in this study to a "null" dataset within XGR via a hypergeometric

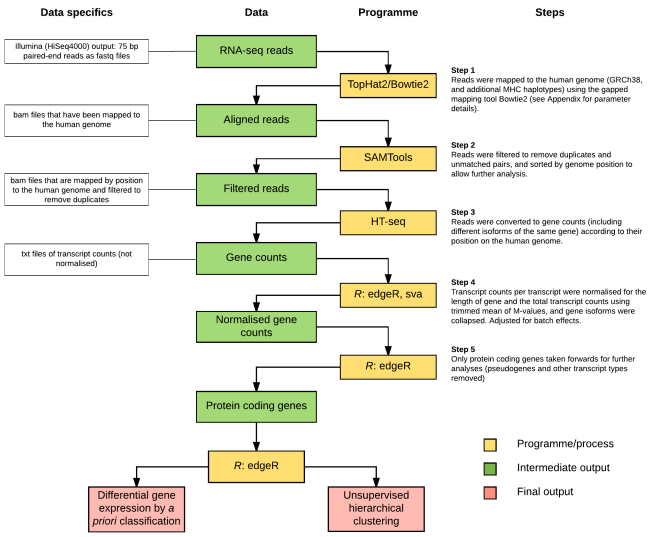


Figure 2.3: Schematic overview of the bioinformatic processing of RNA samples from children enrolled in the study of the transcriptomic response to pneumonia.

distribution function. I used a relatively stringent false discovery rate of 0.01 for gene pathway analysis.

The final analytical steps for (supervised and unsupervised clustering) are described in 6. Figure 2.3 provides an overview of these methods.

Table 2.2: Media used for the separation of PBMCs and the generation of ALS.

Heparinised centrifuge tube	35 μl of heparin 1:1000 (preservative free) in a 15 ml centrifuge tube, stored at 4°C.
Phosphate buffer saline	PBS tablets (Sigma-Aldrich) for reconstitution, 1 tablet to 200 ml of deionised water (followed by sterilisation in an autoclave). Stored at 4°C.
Protease inhibitors	Used at 40X concentration (cOmplete mini EDTA-Free, Roche, Switzerland).
R0 medium	RPMI medium 1640 with the addition of penicillin (500 units/l; Sigma-Aldrich), streptomycin (0.25 mg/l; Sigma-Aldrich) and L-glutamine (10 mmol/l; Sigma-Aldrich). L-glutamine was replenished every 14 days. Stored at 4°C.
R10 medium	R0 medium plus 10% fetal bovine serum (Sigma-Aldrich; heat inactivated and sterile filtered through a 22 μm filter). Stored at 4°C, and discarded if not used within 2 weeks.
RPMI medium 1640 (Hepes modification, without L-glutamine)	Proprietary cell culture medium (Sigma-Aldrich). Stored at 4°C.

3

Childhood invasive pneumococcal disease and pneumonia at Patan Hospital

3.1 Introduction

Severe infections, including pneumonia and invasive bacterial disease (IBD), are the predominant cause of childhood mortality outside the neonatal period worldwide[226]. Pneumonia is the most common syndrome associated with severe infection in children, and the burden is particularly high in South Asia. Prior to the introduction of immunisation schedules containing pneumococcal conjugate vaccination (PCV), *Streptococcus pneumoniae* was the most prevalent pathogenic bacterium isolated from the blood of children with pneumonia across low, middle and high income settings (see Chapter 1[75–78]). However, the sensitivity of blood culture for a bacterial aetiology of pneumonia is estimated to be approximately 1–4% in children when implemented in a routine clinical setting[2, 33], suggesting that using blood culture alone severely underestimates the

true burden of pneumococcal pneumonia. This is thought to be primarily caused by the paucity of bacteria in blood (since bacteria are thought to be limited to lung mucosa in most cases of bacterial pneumonia), and further compounded by treatment of children with antibiotics prior to blood sampling for culture (see Chapter 1).

Despite these limitations, blood culture data are likely to provide important insights into the aetiology of pneumonia in children, and the sensitivity of bacterial pathogens to antibiotics used for the treatment of pneumonia. In addition, the culture of a bacterial pathogen from the blood of a child with pneumonia is presumed to be highly specific for the aetiology of pneumonia. Thus, blood culture represents a "silver standard" test[74] against which proxy markers of pneumococcal pneumonia, and the sensitivity of other diagnostic tests for pneumonia aetiology, can be assessed.

In the absence of accurate diagnostic tests for the aetiology of bacterial pneumonia, "proxy markers" have been developed for patient care and for assessing the impact of vaccination against specific pathogens (notably *Haemophilus influenzae* type b and pneumococci, see Section 1.2.2). This includes the WHO methodology for standardised interpretation of paediatric chest radiographs to assess the presence or absence of endpoint consolidation (synonym: primary endpoint pneumonia or "chest x-ray positive"), infiltrates, or a "normal" chest radiograph[30]. Additionally, differing thresholds of plasma CRP and procalcitonin concentrations (known to be associated with inflammation and bacterial infection) have been used to discriminate between viral and bacterial causes of childhood infection[111], including pneumonia[227]. Among children with clinically-diagnosed pneumonia, the association between radiographic endpoint consolidation and culture of the pneumococcus from blood or pleural fluid is well established[228]. Thus endpoint consolidation is used as a more specific marker of pneumococcal pneumonia, and other bacterial pneumonias, than clinically-diagnosed pneumonia alone.

NP carriage of pneumococci is thought to be a biological prerequisite for the development of pneumococcal pneumonia[50], although not a disease *per se*. NP carriage of pneumococci is not clearly associated with pneumonia[229]. However, pneumococcal serotype-specific differences in "disease potential" have been described and appear to be

relatively stable in time and space: for example serotype 1 pneumococci are an important cause of invasive pneumococcal disease with a NP carriage prevalence considerably greater in children with pneumonia than healthy children[65–67] (see Chapter 1). NP carriage of pneumococcal serotypes with "high" disease potential, such as serotype 1 pneumococci, could be used to support a diagnosis of pneumococcal pneumonia in combination with other proxy markers of the disease, particularly in the context of epidemiological studies.

In the absence of accurate diagnostic tests for pneumococcal pneumonia, I investigated the clinical epidemiology of invasive pneumococcal disease and pneumonia in children admitted to Patan Hospital, Kathmandu Valley, Nepal (see Section 2.1). I aimed to describe the burden of these diseases in children at Patan Hospital, and to develop proxy markers of pneumococcal pneumonia that could be used in further analyses of 10-valent PCV impact on childhood pneumonia (Chapter 4), and for the assessment of novel diagnostic tests for pneumococcal pneumonia (Chapters 5 and 6).

Here, I initially describe the burden of community-acquired IBD in neonates, infants and children <14 years of age admitted to Patan Hospital. I use data from the childhood IBD surveillance programme at Patan Hospital, this includes culture of blood, CSF (in cases of suspected meningitis) and pleural fluid (in a small number of children with pneumonia requiring pleurocentesis) from March 2005 to end of December 2016. I focus my analysis on invasive pneumococcal disease. During the period of surveillance, the Nepal infant immunisation schedule incorporated Bacille-Calmette-Guérin (BCG), diphtheria-tetanus-pertussis-hepatitis B, oral polio vaccine and measles vaccine. Vaccination against *H. influenzae* type b was introduced incrementally during 2009, inactivated polio vaccination in 2014 and 10-valent pneumococcal conjugate vaccination in 2015. Vaccine uptake was high across Nepal (see Chapter 1[207]).

Following this, I describe the characteristics of children, aged 2 months to 14 years of age, with clinician defined pneumonia admitted to Patan Hospital from March 2014 until the end of December 2016. This study of childhood pneumonia (the Inpatient Pneumonia Study) was conducted in the context of the introduction of 10-valent PCV to the infant

immunisation schedule on 21st August 2015 (in the Kathmandu Valley). 10-valent PCV was implemented with a unique Nepali schedule at 6 weeks, 10 weeks and 9 months of age. There was a limited "catch-up campaign" of one or two doses of 10-valent PCV in children aged 10 weeks to <1 year.

The primary aim of the Inpatient Pneumonia Study was to assess the impact of 10-valent PCV on proxy markers of pneumococcal pneumonia. Preliminary data to address this aim are described in Chapter 4.

3.2 Methods

3.2.1 Overview of studies

Studies of childhood infection, including continuous surveillance of IBD in children admitted to Patan Hospital, have been undertaken since 2005[4, 6, 52, 68, 230, 231]. Figure 3.1 shows an overview of the studies in Patan Hospital. Detailed data from 2005–2006 have previously been published[4, 68], as have data from 2007–2009[6]. Data from 2010–2016 have not previously been published.

Data from the surveillance programme and the Inpatient Pneumonia Study were analysed in Stata 14 (Statacorp, USA) and R[218]. Data underwent extensive quality checks, and queries in the data were cross-checked with original records from the Microbiology Laboratory at Patan Hospital.

Ethics statement

Informed consent from caregivers was not required for enrolment of children to the IBD surveillance programme at Patan Hospital on the basis of anonymity and no deviation from standard clinical care (Nepal Health Research Council, NHRC, reference 517 and Oxford Tropical Research Ethics Committee, OXTREC, 032-06). For enrolment to the Inpatient Pneumonia Study, caregivers gave informed, written consent. The NHRC (reference 04/14) and OXTREC (reference 05-14) agreed the study protocol.

Statistical methods Descriptive statistics were prepared using Stata v.14.1 (Statacorp, TX, USA) and R[218]. Pearson's χ^2 test without continuity corrections was used to compare two or three groups for categorical variables; Student's t-test or one-way analysis of variance (ANOVA) was used to compare continuous variables, following $\log_10$ transformation to normalise data if appropriate.

3.2.2 Invasive pneumococcal disease, April 2005–December 2016

All children <14 years of age with suspected community-acquired IBD admitted to Patan Hospital were included in the surveillance programme. Suspected IBD was defined as suspected pneumonia, suspected meningitis or suspected sepsis at admission, on the basis of internationally agreed guidelines[94], but with an element of clinician discretion. In practice, 77% of all admitted children were defined as suspected IBD (Figure 3.1).

Blood was taken aseptically on all children included in IBD surveillance for culture using Bactec Paeds Plus bottles (Becton Dickinson) and methods detailed in Section 2.2.2. CSF was sampled from children with suspected meningitis unless contra-indicated, and cultured using standard methods[214]. Pleural fluid was cultured from children who required pleurocentesis, and cultured using standard methods[214]. Serotyping was by the Quellung reaction, undertaken at the WHO reference laboratory, Christian Medical College, Vellore, India, with additional serotyping and quality control as described in Section 2.3.2.

Between 2005 and end of December 2006, data on all children <14 years of age with suspected IBD were available. Between 2007 and July 2009, data on children with positive cultures from blood, CSF and pleural fluid were available since only limited childhood IBD surveillance was undertaken during this period. For clarity I report data from 2005 to end of December 2009 as numbers of culture positive cases, and from 2010 to end of December 2016 as both numbers of culture positive cases, and proportions of all children admitted with suspected IBD. Data from were entered on specific databases held at Patan Hospital for each year of surveillance, and integrated to form one continuous

dataset, with cross-checking of positive culture results with a separate database held in the Department of Microbiology at Patan Hospital.

3.2.3 Childhood pneumonia at Patan Hospital, March 2014–December 2016

A detailed study of pneumonia in children admitted to Patan Hospital (the Inpatient Pneumonia Study) has been undertaken since 20th March 2014. Here, I present data to the end of December 2016. Eligibility criteria were: aged between 2 months and 14 years, admitted to Patan Hospital, with a clinical diagnosis of pneumonia, and caregiver consent for enrolment to the study. Inborn neonates who developed pneumonia prior to discharge from hospital were not eligible. Children who were enrolled, discharged, and then readmitted with pneumonia within one month of discharge were not re-enrolled for the second admission. The diagnosis of pneumonia was made by the admitting paediatric team who reviewed all patients prior to a decision to admit to hospital. Diagnoses of pneumonia were based on standard clinical criteria in children <5 years of age[15], and with a degree of clinical discretion in young and older children. Chest radiographs were not normally available before a decision was made to admit a patient with pneumonia. Data from the initial presentation to hospital were recorded on study-specific case record forms. A member of the clinical research team was based on the Paediatric Ward of Patan Hospital from 08:00 h to 20:00 h every day (including weekends) and reviewed all admissions for eligibility.

Digital chest radiographs were taken at, or on the day following, admission to hospital on enrolled children and were assessed as previously described by a paediatrician and radiologist with arbitration of discordant results settled by a second radiologist (Section 2.3). In brief, two readers (MJC and a radiologist) interpreted chest radiographs as "endpoint consolidation", "infiltrates" or "normal", with arbitration of disagreements by a third reader (a senior radiologist) according to the standardised WHO-methodology[30].

NP samples were taken on all children enrolled into the Inpatient Pneumonia Study, using a single nylon flocked swab (Thermo Fisher Scientific). Swabs were placed immediately into STGG media, taken to the Microbiology Laboratory at Patan Hospital and cultured

for pneumococci as described in Section 2.3.2. Serotyping was via the Quellung reaction at Patan Hospital with extensive external quality control.

Patients enrolled into the Inpatient Pneumonia Study were all included in the childhood IBD surveillance programme. Blood, and where indicated, pleural fluid and CSF, was aseptically sampled and processed as previously described (Section 2.2.2). Blood neutrophil counts were measured from additional blood samples taken at admission using standard methods at Patan Hospital. From February 2015 onwards, plasma CRP concentrations were retrospectively measured in plasma on enrolled children.

The clinical, laboratory and microbiological characteristics used to develop proxy markers of pneumococcal pneumonia in this study of childhood pneumonia are: the components of the WHO clinical classification of pneumonia ("fast breathing", "chest indrawing" and "danger signs"[23]), absence of wheeze on examination, endpoint consolidation on chest radiographs[30], inflammatory markers (blood neutrophil count and plasma CRP concentrations), and data on the NP carriage of pneumococcal serotypes[67]. Absence of wheeze on examination has previously shown to be associated with pneumococcal pneumonia in this setting[4]), hence its inclusion as a proxy marker of pneumococcal pneumonia here. The WHO clinical classification is only applied to children <5 years of age, and is therefore not used for descriptions of the entire case series.

3.3 Results

3.3.1 Invasive pneumococcal disease, April 2005–December 2016

Enrolment to the childhood IBD surveillance programme

Between March 2005 and December 2016, data from 13760 admissions were available for analysis. Between March 2005 and December 2009 (where only a proportion of children with suspected IBD were enrolled), 3375 children were enrolled. Between 2010 and December 2016, 10385 children were enrolled to the surveillance programme (Figure 3.1). During the period 2010–2016 the median age of children enrolled to the IBD surveillance programme was 0.9 years (IQR 0.1, 3.1), and 61% of children were male.

The numbers and demographic characteristics of all enrolled children are shown in Table 3.1.

Childhood IBD at Patan Hospital

Of 10385 children enrolled between January 2010 and December 2016, the most prevalent pathogen isolated from normally sterile sites was *S. enterica* var Typhi, in 91 (0.9%) children. *S. enterica* var Paratyphi was identified in an additional 17 (0.2%) children. During this period, *S. pneumoniae* was the second-most prevalent cause of IBD, isolated in 59 (0.6%) children. Table 3.2 shows the number and proportion of children with IBD enrolled into the surveillance programme during this period. Figure 3.2 illustrates the number of isolates from children during the surveillance programme from March 2005 to December 2016.

Childhood invasive pneumococcal disease at Patan Hospital

Across all years of the surveillance programme (i.e. March 2005–December 2016), *S. pneumonie* was isolated from normally sterile sites in 128 children. Of these 128 children with invasive pneumococcal disease, 6 (5%) children had pneumococci isolated from pleural fluid only and 3 (2%) children had pneumococci isolated from CSF only – all other children had pneumococcal bacteraemia.

S. pneumoniae was isolated in all years of the IBD surveillance programme, although there was an apparent decline in the annual number of isolates between the period 2005–2009 and 2010–2016. It was not possible to formally analyse this decline since data were not available for children admitted to Patan Hospital without (proven) invasive bacterial (including pneumococcal) disease during 2005–2009. This was most pronounced in the youngest children and was not anecdotally associated with a change in blood sampling or microbiological culture methods (Figure 3.3). The majority (67%) of invasive pneumococcal disease was identified in children $\geq$2 years of age – with 38% in children $\geq$5 years of age; 61% was in male children. Invasive pneumococcal disease was not diagnosed in neonates (<1 month of age) across the surveillance period. Table 3.1 details the demographic characteristics of children with invasive pneumococcal disease across all years. As noted above, data were not available to describe changes in the

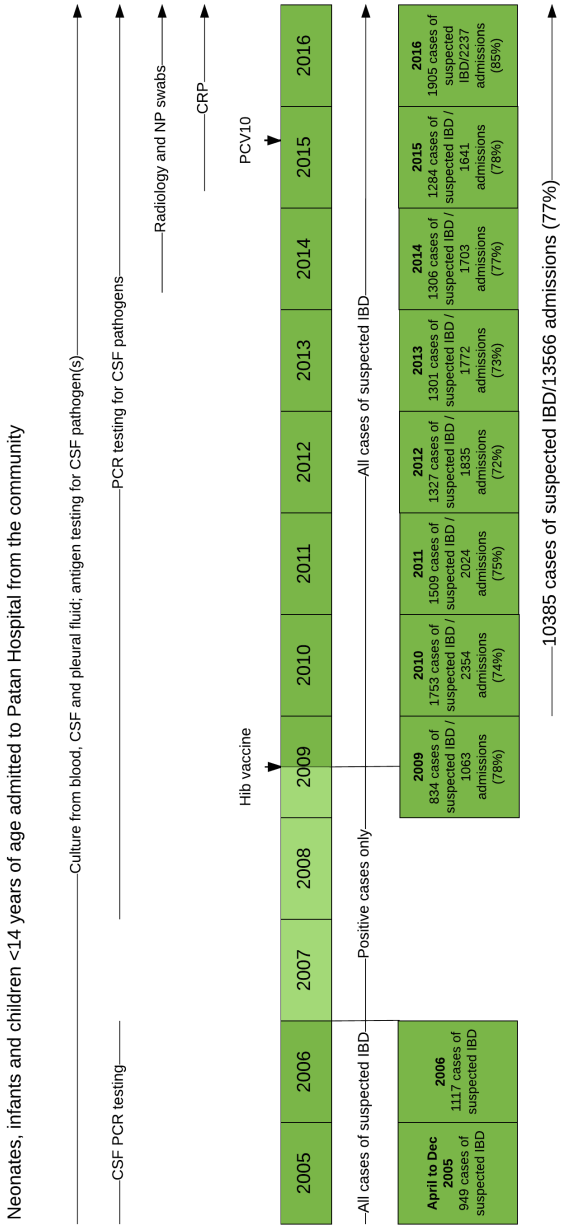


Figure 3.1: Enrolment of children 0–14 years of age with community-acquired disease to the childhood IBD surveillance programme at Patan Hospital by year. Key microbiological methods are described as horizontal lines above the figure (only culture results from blood, CSF and pleural fluid are described here), and vaccine introduction by arrows. Darker green years represent enrolment of all children admitted with suspected IBD, lighter green cases represent enrolment of a smaller subset of these cases. For clarity, the data were separated into a period 2005–2009, and 2010–2016 as described in the text. Hib, *H. influenzae* type b; PCV10, 10-valent PCV.

Table 3.1: Demographic characteristics of children enrolled into the childhood IBD surveillance programme, and of children with invasive pneumococcal disease.

Year	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**All years**
Total admissions	-	-	-	-	-	2354	2024	1835	1772	1703	1641	2237	13566
Suspected IBD	-	-	-	-	-	1753 (74%)	1509 (75%)	1327 (72%)	1301 (73%)	1306 (77%)	1284 (78%)	1905 (85%)	10385 (77%)
Male sex	-	-	-	-	-	1019 (58%)	930 (62%)	803 (61%)	796 (61%)	807 (62%)	809 (63%)	1155 (61%)	6319 (61%)
Age (years; median, IQR)	-	-	-	-	-	1.0 (0.1, 3.6)	0.8 (0.1, 2.8)	0.8 (0.1, 2.4)	1.0 (0.1, 2.9)	1.0 (0.1, 3.1)	1.0 (0.1, 3.0)	1.1 (0.2, 3.4)	0.9 (0.1, 3.1)
< 1 month	-	-	-	-	-	350 (20%)	294 (20%)	302 (23%)	274 (21%)	282 (22%)	258 (20%)	312 (16%)	2072 (20%)
1 month to <6 months	-	-	-	-	-	314 (18%)	305 (20%)	232 (18%)	259 (20%)	227 (17%)	237 (18%)	338 (18%)	1912 (18%)
6 months to <1 year	-	-	-	-	-	213 (12%)	213 (14%)	172 (13%)	153 (12%)	140 (11%)	149 (12%)	253 (13%)	1293 (12%)
1 year to <2 years	-	-	-	-	-	271 (15%)	200 (13%)	205 (16%)	196 (15%)	205 (16%)	197 (15%)	287 (15%)	1561 (15%)
2 years to <5 years	-	-	-	-	-	266 (15%)	256 (17%)	185 (14%)	221 (17%)	227 (17%)	229 (18%)	377 (20%)	1761 (17%)
5 years and older	-	-	-	-	-	339 (19%)	230 (15%)	213 (16%)	197 (15%)	220 (17%)	212 (17%)	334 (18%)	1745 (17%)
Year	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**All years**
Total with IPD[a]	16	9	16	17	11	8	7	10	7	10	8	9	128
Male sex	8	5	11	9	8	8	5	3	4	4	4	7	78 (61%)
Age (years; median IQR)													3.9 (1.2, 6.3)
< 1 month	0	0	0	0	0	0	0	0	0	0	0	0	0
1 month to <6 months	5	2	3	1	0	0	0	1	0	0	0	0	12 (10%)
6 months to <1 year	1	1	3	1	1	2	1	4	0	2	1	0	17 (13%)
1 year to <2 years	3	2	1	2	1	1	0	0	1	0	1	1	13 (10%)
2 years to <5 years	2	2	3	7	6	2	3	2	4	3	2	2	38 (30%)
5 years and older	5	2	6	6	3	3	3	3	2	5	4	6	48 (38%)

[a]IPD, invasive pneumococcal disease.

Table 3.2: Number (and percentage) of children admitted to Patan Hospital with suspected IBD, with bacteria isolated from blood, CSF or pleural fluid, by year 2010–2016. Strep., streptococci; staph., staphylococci.

	Number of isolates (and percentage of all enrolled)							
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
Gram-positive organisms								
β-haemolytic strep. (other)	1 (0.1%)	0	1 (0.1%)	0	0	0	0	2 (<0.1%)
Enterococcus spp.	1 (0.1%)	0	0	0	0	0	0	1 (<0.1%)
Group A β-haemolytic strep.	0	0	0	0	2 (0.2%)	1 (0.1%)	1 (0.1%)	4 (<0.1%)
Staphylococcus aureus	3 (0.2%)	3 (0.2%)	3 (0.3%)	3 (0.3%)	4 (0.3%)	2 (0.2%)	14 (0.7%)	32 (0.3%)
Streptococcus pneumoniae	8 (0.5%)	7 (0.5%)	10 (0.8%)	7 (0.6%)	10 (0.8%)	8 (0.6%)	9 (0.5%)	59 (0.6%)
Viridans streptococci	5 (0.3%)	3 (0.2%)	4 (0.4%)	3 (0.3%)	6 (0.5%)	1 (0.1%)	11 (0.6%)	33 (0.3%)
Gram-negative organisms								
Acinetobacter spp.	7 (0.4%)	2 (0.2%)	3 (0.3%)	2 (0.2%)	4 (0.3%)	2 (0.2%)	9 (0.5%)	29 (0.3%)
Enterobacter spp.	1 (0.1%)	1 (0.1%)	1 (0.1%)	6 (0.5%)	3 (0.2%)	2 (0.2%)	1 (0.1%)	15 (0.1%)
Escherichia coli	2 (0.2%)	1 (0.1%)	2 (0.2%)	1 (0.1%)	1 (0.1%)	2 (0.2%)	5 (0.3%)	14 (0.1%)
Haemophilus influenzae	0	0	1	0	0	0	0	1 (<0.1%)
Klebsiella oxytoca	0	0	0	0	0	0	1 (0.1%)	1 (<0.1%)
Klebsiella pneumoniae	0	0	3 (0.3%)	3 (0.3%)	5 (0.4%)	1 (0.1%)	4 (0.2%)	16 (0.2%)
Klebsiella spp.	2 (0.2%)	1 (0.1%)	0	0	1 (0.1%)	0	0	4 (<0.1%)
Neisseria meningitidis	1 (0.1%)	0	1 (0.1%)	0	0	0	1 (0.1%)	3 (<0.1%)
Neisseria spp.	2 (0.2%)	0	0	0	0	0	0	2 (<0.1%)
Pseudomonas aeruginosa	0	0	0	0	0	0	3 (0.2%)	3 (<0.1%)
Pseudomonas spp.	1 (0.1%)	1 (0.1%)	0	0	1 (0.1%)	0	2 (0.1%)	5 (<0.1%)
S. enterica var Paratyphi A	6 (0.4%)	5 (0.4%)	2 (0.2%)	0	2 (0.2%)	1 (0.1%)	1 (0.1%)	17 (0.2%)
S. enterica var Typhi	20 (1.2%)	9 (0.6%)	14 (1.1%)	22 (1.7%)	11 (0.8%)	7 (0.5%)	8 (0.4%)	91 (0.9%)
Salmonella spp.	3 (0.2%)	1 (0.1%)	1 (0.1%)	0	0	0	0	5 (<0.1%)
Serratia marcescens	0	0	0	0	0	1 (0.1%)	0	1 (<0.1%)
Other								
Yeast	1 (0.1%)	1 (0.1%)	1 (0.1%)	1 (0.1%)	0 (0%)	2 (0.2%)	0 (0%)	6 (0.1%)
Total pathogens	64 (3.7%)	35 (2.4%)	47 (3.6%)	48 (3.7%)	50 (3.8%)	30 (2.3%)	70 (3.7%)	344 (3.3%)
Not done	0 (0%)	0 (0%)	39 (3%)	160 (12.3%)	31 (2.4%)	42 (3.3%)	98 (5.1%)	370 (3.6%)
Negative	1549 (88.4%)	1374 (91.1%)	1169 (88.1%)	1043 (80.2%)	1165 (89.2%)	1178 (91.7%)	1618 (84.9%)	9096 (87.6%)
Uncertain significance								
Coagulase-neg. staph. (neonates)	16	27	8	12	16	8	16	103
Mixed: *Serratia*, *Klebsiella* spp.	1	0	0	0	0	0	0	1
Contaminants								
Bacillus spp.	41	0	0	6	10	7	8	72
Coagulase-neg. staph.	61	21	5	12	20	4	38	161
Micrococcus spp.	21	6	0	4	13	7	12	63
Unknown contaminant	0	46	59	16	1	8	45	175
Total	**1753**	**1509**	**1327**	**1301**	**1306**	**1284**	**1905**	**10385**

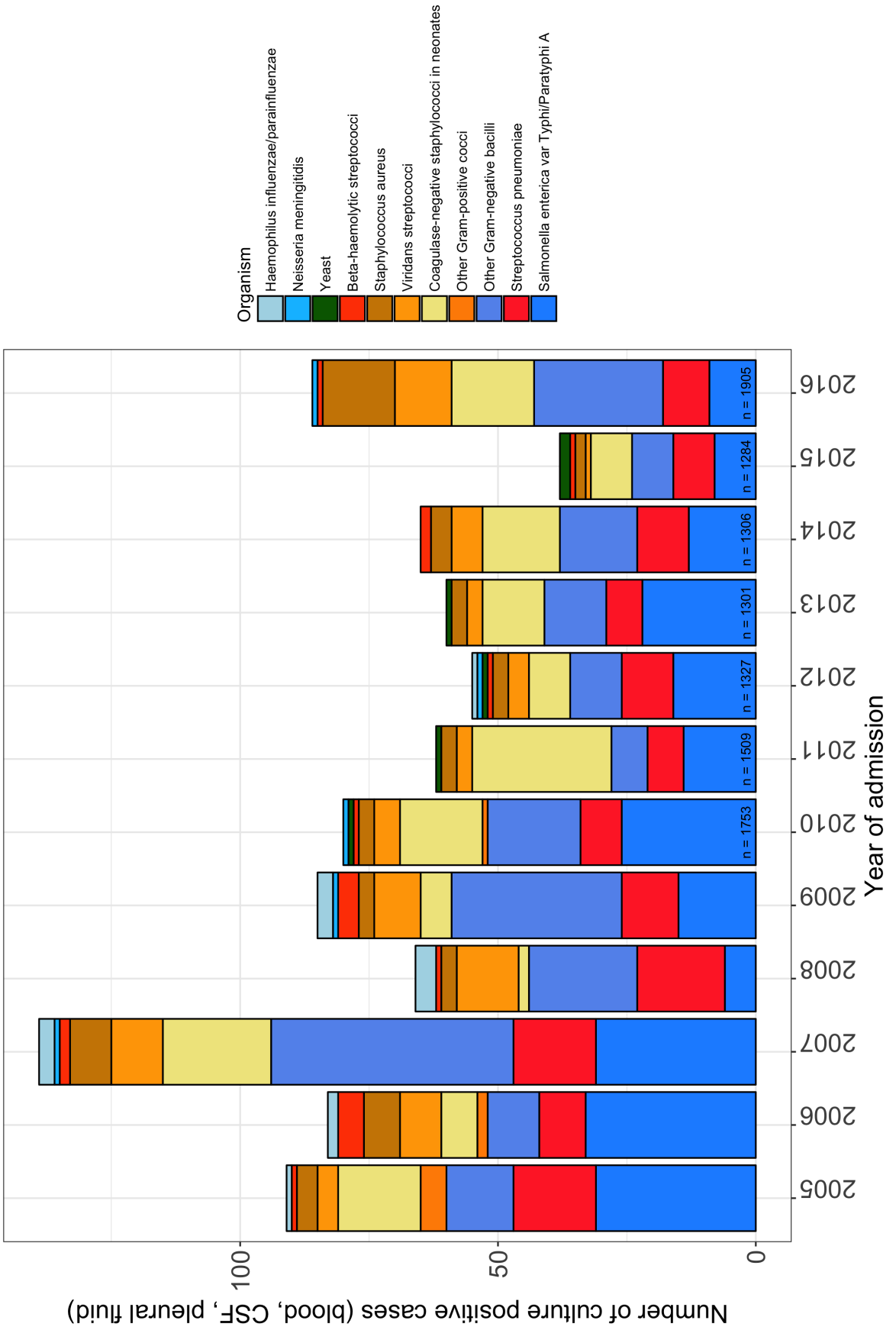


Figure 3.2: Childhood IBD at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures by organism/group by year.

age structure of all children entering surveillance for suspected IBD during the period 2005–2009.

Serotype data were available on 124/128 (97%) pneumococcal isolates from invasive pneumococcal disease. The most commonly isolated pneumococcal serotype was serotype 1 (58, 47% of isolates), followed by serotype 5 (14, 11% of isolates) and serotype 14 (6, 5% of isolates). Of invasive pneumococcal disease caused by serotype 1 pneumococci, 6 (10%) isolates were from children <2 years of age, 22/58 (38%) isolates were from children 2 years to 5 years of age and 30/58 (51%) were from children $\geq$5 years of age. The median age of children with invasive pneumococcal disease caused by serotype 1 was 5.1 years (IQR 3.9, 7.6) in comparison with 1.8 years (IQR 0.67, 5.2; two-sided t-test, $p < 0.001$) for children with invasive pneumococcal disease caused by other serotypes.

The age distribution of pneumococcal isolates by their presence in 10-valent and 13-valent PCV is shown in Figure 3.4, and the age distribution of serotype 1 pneumococcal isolates is shown in Figure 3.5. Isolates from children <2 years of age were not significantly less likely to of serotypes included in 10-valent PCV (22/42, 52%) than children aged 2 years to <5 years (29/38, 76%) and those aged $\geq$5 years (42/48, 88%, χ^2 test, p = 0.117). There was, however, no trend towards an increase or decrease in the proportion of pneumococcal serotypes included in the 10-valent PCV (in these almost entirely pre-vaccine data) by year (Figure 3.6).

3.3.2 Childhood pneumonia at Patan Hospital, March 2014–December 2016:

Enrolment of children to the Inpatient Pneumonia Study at Patan Hospital

There were 5125 episodes of children admitted to Patan Hospital from the community (excluding surgical cases) during the period March 2014 to the end of December 2016. Of these, 938 (18%) children were clinically diagnosed with pneumonia and enrolled into this study. Of children enrolled to the study, 483 (51%) had received prior treatment with antibiotics. Consent for enrolment was not given in an unrecorded, but small number of children (<5% of eligible children). No cases of children clinically diagnosed with

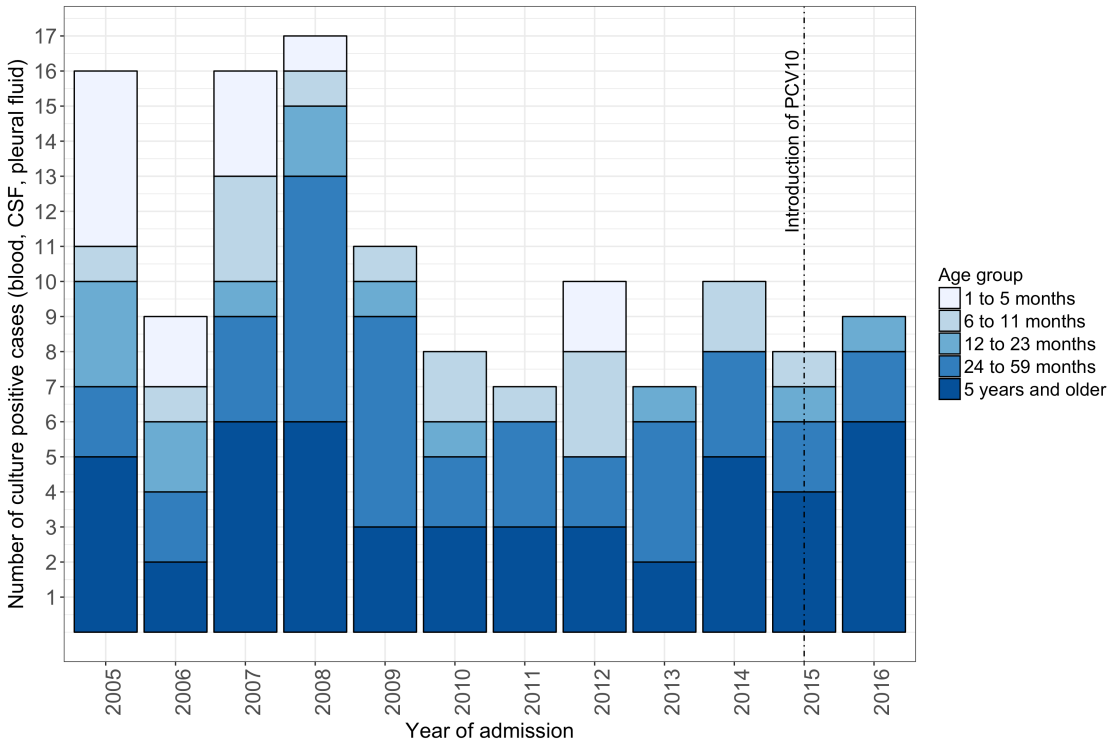


Figure 3.3: Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures for *Streptococcus pneumoniae* by age group of child year of admission.

pneumonia were otherwise missed. Audit of enrolment (data not shown) suggests that a larger proportion of admitted children would have been diagnosed with pneumonia is the WHO clinical criteria for pneumonia were used as inclusion criteria (see 1 for a discussion of this).

Characteristics of children in the Inpatient Pneumonia Study: age group

The median age of children enrolled to the study was 0.9 years (IQR 0.5, 1.8). Of 938 children enrolled, 611 (65%) children were <2 years of age, 215 (23%) children were 2 to 5 years of age and 112 (12%) children were 5 to 14 years of age. The group of children <5 years of age are referred to as "young children" hereafter. The WHO clinical classification of pneumonia[15] is applicable to children 2 months to <5 years of age only.

Detailed characteristics of patients enrolled to the study, in total and by age group are shown in Table 3.3. There was no difference in the proportion of young children with

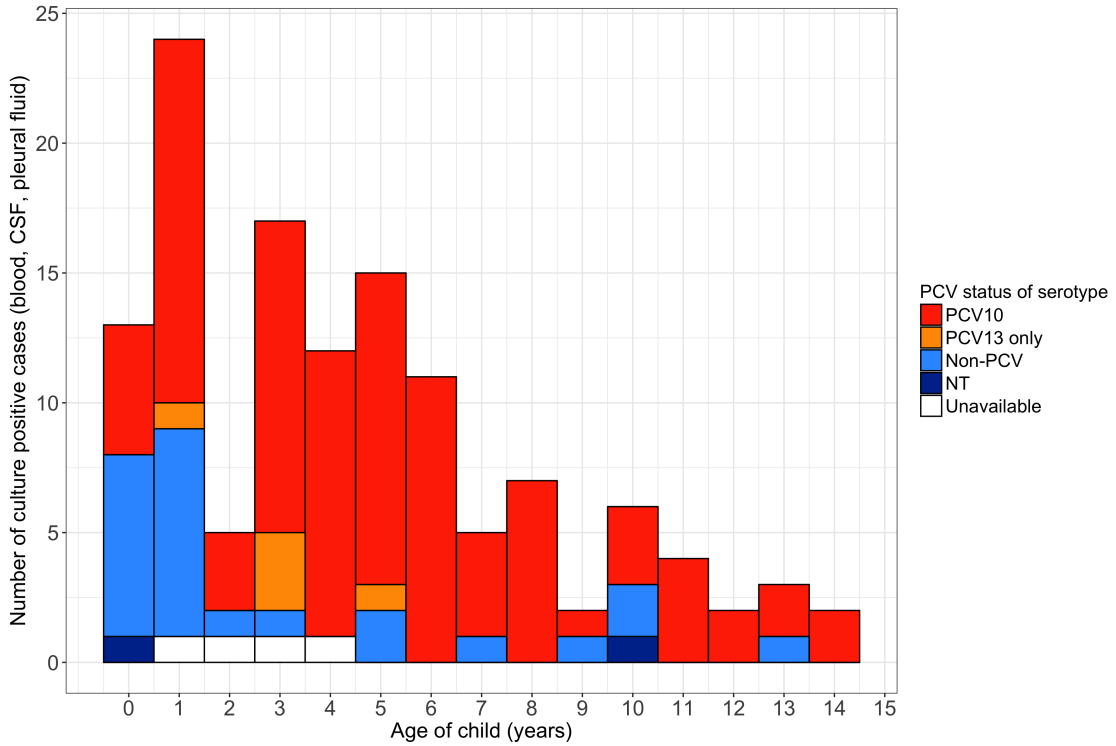


Figure 3.4: Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures for *Streptococcus pneumoniae* by age of child and presence of serotype in 10-valent and 13-valent PCV. PCV10, 10-valent PCV; PCV13 only, serotypes 3, 6A and 19A additionally contained in 13-valent PCV.

fast breathing pneumonia between <2 years of age and those 2 to 5 years of age; but children <2 years of age were more likely to have chest indrawing 161/611 (26%) than in children 2 to 5 years of age (137/215, 17%, p = 0.006). Older children (those $\geq$5 years of age) were more likely to not have wheeze on examination (98/112, 88%) than children 2 to 5 years of age (140/215, 65%) and children <2 years of age (291/611, 48%, p < 0.001). Data on haemoglobin oxygen saturations were not routinely recorded for research purposes.

Increased neutrophil count was associated with membership of older age groups (median 6.0 x 10^3/mm^3 in children <2 years of age vs 8.1 x 10^3/mm^3 in children 2 to 5 years of age vs 12 x 10^3/mm^3 in children 5 to 14 years of age, ANOVA, p < 0.001). In the 474 children who had plasma CRP concentration measured, higher CRP concentration was associated with older age groups (median 24 mg/l in children <2 years of age vs 33 mg/l in children 2 to 5 years of age vs 102 mg/l in children 5 to 14 years of age, ANOVA, p <

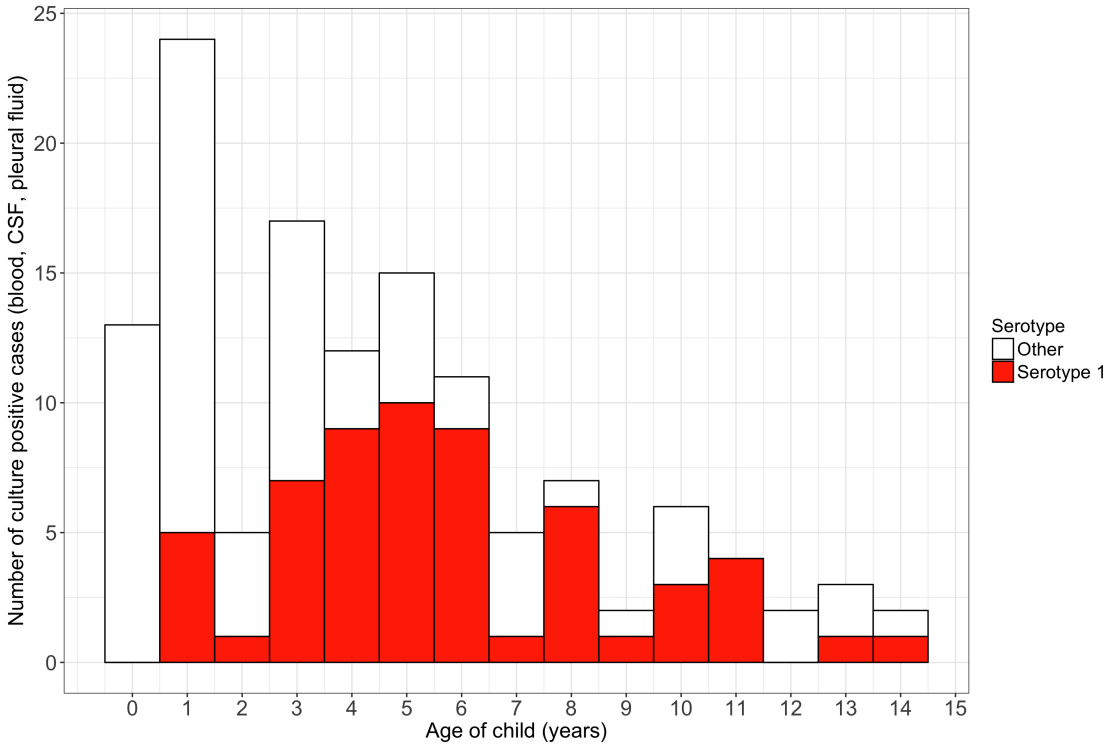


Figure 3.5: Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: numbers of positive blood, CSF or pleural fluid cultures for *Streptococcus pneumoniae* by age of child and whether the isolate was of serotype 1.

0.001). Reference ranges for neutrophil counts and CRP concentrations are not available for this population.

In total, 326/938 (35%) children enrolled had radiographic endpoint consolidation diagnosed. Radiographic endpoint consolidation was associated with older age group: diagnosed in 149/611 (24%) children <2 years of age, 100/215 (47%) children 2 years to <5 years of age, and 77/112 (69%) children 5 years to 14 years of age ($p < 0.001$). In contrast, radiographic infiltrates (without endpoint consolidation) were associated with younger age group: diagnosed in 91/611 (15%) children <2 years of age, 23/215 (11%) children 2 years to <5 years of age, and 8/112 (6.9%) children 5 years to 14 years of age ($p = 0.04$).

Antibiotic therapy prior to NP sampling, carriage of any pneumococcal serotype, and carriage of non-vaccine-type pneumococci (either additional PCV13 serotypes, non-PCV serotype or non-typeable) were not associated with age. In contrast, NP carriage of vaccine-type pneumococci was associated with older age group: cultured in 69/611

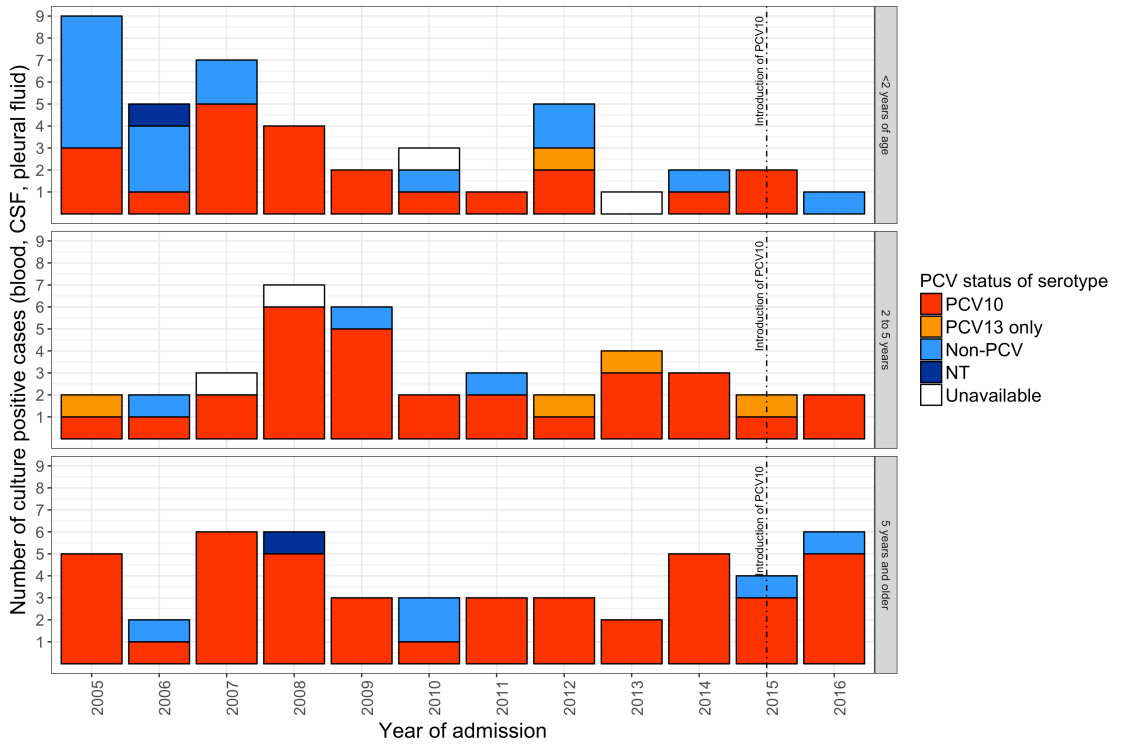


Figure 3.6: Childhood invasive pneumococcal disease at Patan Hospital 2005–2016: relationship between pneumococcal serotype and presence in 10-valent and 13-valent PCV by year and by age group. PCV10, 10-valent PCV; PCV13 only, serotypes 3, 6A and 19A additionally contained in 13-valent PCV.

(11%) children <2 years of age, 28/215 (13%) children 2 to 5 years of age, and 26/112 (23%) children 5 to 14 years of age (χ^2 test of trend, p = 0.003).

Pneumococci were cultured from blood in 15 children and from pleural fluid in an additional 2 children – totaling 17/938 (1.8%) children with pneumonia. Invasive pneumococcal disease was associated with older age group: in 2/611 (0.3%) children <2 years of age, 4/215 (1.9%) children 2 to 5 years of age, and 11/112 (9.8%) children 5 to 14 years of age (χ^2 test of trend, p < 0.001).

Characteristics of children in the Inpatient Pneumonia Study: absence of wheeze

Of 938 children enrolled to the study, 529 (55%) did not have wheeze on their admission clinical examination. Children without wheeze on examination were older (median age 1.7 years vs 0.9 years, chi^2 test, p < 0.001) than those with wheeze. A higher proportion of children without wheeze had general danger signs in comparison to children with wheeze (148/529, 28% vs 78/409, 19%, chi^2 test, p = 0.002). Children without wheeze

Table 3.3: Characteristics of patients enrolled in the study by age group.

	Age group				
	2 months to <2 years	2 to <5 years	5 to 14 years	All children	p[a]
Total	611 (65%)	215 (23%)	112 (12%)	938	
Male sex	380 (62%)	127 (59%)	74 (66%)	581 (62%)	0.454
Clinical features[b]					
Fast breathing pneumonia	454 (74%)	149 (69%)	–	603 (73%)	0.183
Chest indrawing pneumonia	161 (26%)	36 (17%)	–	197 (24%)	0.006
Chest indrawing or fast breathing pneumonia	469 (77%)	153 (71%)	–	622 (75%)	0.122
General danger signs[c]	144 (24%)	60 (28%)	–	226 (27%)	0.239
No wheeze on examination	291 (48%)	140 (65%)	98 (88%)	529 (56%)	< 0.001
Laboratory results					
CRP mg/l (median, IQR)[d] <2 years, 2 to <5 years and 5 to 14 years	24 (2.5, 27)	33 (11, 95)	102 (23, 205)	15 (4, 52)	< 0.001
Neutrophil count x 10^3/mm^3 (median, IQR)	6.0 (3.7, 9.5)	8.1 (4.8, 13.6)	12 (7.0, 19)	7.0 (4.0, 11)	< 0.001
Radiographic results					
Endpoint consolidation	149 (24%)	100 (47%)	77 (69%)	326 (35%)	< 0.001
Infiltrates only	91 (15%)	23 (11%)	8 (6.9%)	122 (13%)	0.04
No EPC, uninterpretable for infiltrates	9 (1.4%)	1 (0.5%)	0	10 (1.1%)	
Normal	338 (55%)	81 (38%)	20 (18%)	439 (47%)	
Uninterpretable, not done	24 (3.9%)	10 (4.7%)	1 (0.9%)	35 (3.7%)	
NP carriage of pneumococcal serotypes					
Prior antibiotic therapy	306 (50%)	115 (53%)	62 (55%)	483 (51%)	0.47
Carriage of any pneumococci	210 (34%)	74 (34%)	36 (32%)	320 (34%)	0.900
Carriage of any pneumococci if no prior antibiotics	141/301 (47%)	49/99 (49%)	28/49 (57%)	218/449 (49%)	0.400
10-valent PCV (VT) pneumococci	69 (11%)	28 (13%)	26 (23%)	123 (13%)	0.003
Additional 13-valent PCV pneumococci	23 (3.8%)	7 (3.3%)	1 (0.9%)	31 (3.3%)	0.295
Non-PCV pneumococci	83 (14%)	30 (14%)	7 (6.3%)	120 (13%)	0.086
Non-typeable pneumococci	35 (5.7%)	9 (4.2%)	2 (1.8%)	46 (4.9%)	0.177
IBD					
Invasive pneumococcal disease[e]	2 (0.3%)	4 (1.9%)	11 (9.8%)	17 (1.8%)	< 0.001
Other invasive pathogens[f]	8 (1.3%)	1 (0.5%)	3 (2.6%)	12 (1.3%)	0.238

[a]p values calculated using χ^2 test without continuity corrections for two or three groups as necessary, or using one-way analysis of variance (ANOVA) of $\log_{10}$ transformed continuous variables (CRP and blood neutrophil count).

[b]825/938 cases complete for fast breathing and chest indrawing.

[c]Danger signs: unable to drink, persistent vomiting, convulsions, lethargy or reduced consciousness, stridor in a calm child or severe malnutrition (weight-for-age z-score <-3).

[d]Measured in 474 children in age groups <2 years, 2 to <5 years and 5 to 14 years respectively: n = 300, n = 115, n = 59

[e]isolated from blood or pleural fluid

[f]In age groups <2 years, 2 to <5 years and 5 to 14 years respectively: *Acinetobacter* spp. 2, 0, 1; *Escherichia coli* 1, 0, 0; *Neisseria meningitidis* 1, 0, 0; *Pseudomonas* spp. 1, 1, 0; *Salmonella enterica* var Typhi 1, 0, 0; *Staphylococcus aureus* 1, 0, 2; viridans streptococci 1, 0, 0.

also had higher neutrophil counts (median 5.9 x 10^3/mm^3 vs 7.9 x 10^3/mm^3, ANOVA, p < 0.001) and higher plasma CRP concentrations (median 28 mg/l vs 9.1 mg/l, ANOVA, p < 0.001).

Of the 529 children who did not have wheeze on examination, 238 (45%) were diagnosed with radiographic endpoint consolidation in comparison to 88/409 (22%) children with wheeze (p < 0.001). A lower proportion of children without wheeze were diagnosed with radiographic infiltrates (58/529, 11% vs 64/409, 16%, chi^2 test, p = 0.044). A slightly higher proportion of children without wheeze received antibiotics (either by parental report or in hospital) prior to NP sampling in comparison to children with wheeze (288/529, 54% vs 195/409, 48%, chi^2 test, p = 0.047). However, the proportion of children with NP carriage of any pneumococci was higher in children without wheeze (197/529, 37%, vs 123/409, 30%, chi^2 test, p = 0.026), as was the proportion of children with NP carriage of VT pneumococci (91/529, 17% vs 32/409, 7.8%, chi^2 test, p < 0.001) in comparison to children with wheeze. Invasive pneumococcal disease was diagnosed in 16/529 (3.0%) children without wheeze in comparison to 1/409 (0.2%) child with wheeze (chi^2 test, p < 0.001). The single child with invasive pneumococcal disease and wheeze was 8 months of age and also had radiographic endpoint consolidation and a blood neutrophil count of 24 x 10^3/mm^3. Detailed characteristics of patients enrolled to the study, in total and by absence or presence of wheeze are shown in Table 3.4.

Characteristics of children in the Inpatient Pneumonia Study: chest radiographs

Admission chest radiographs were done and interpretable for 887/938 (95%) children. There was a high level of agreement between reader 1 and reader 2 for the diagnosis of endpoint consolidation on chest radiographs in children enrolled to the study (κ = 0.69). Of the chest radiographs reviewed by the arbiter (radiographs with discordant readings, or a random 10%), there was a higher level of agreement with reader 2 (κ = 0.71) than with reader 1 (κ = 0.57). Table 3.5 details the concordance between readers further.

Endpoint consolidation was diagnosed on 326 (37%) radiographs, infiltrates (without endpoint consolidation) were diagnosed on 122 (14%) radiographs and 439 (49%) radiographs were deemed normal. A higher proportion of radiographs with infiltrates were

Table 3.4: Characteristics of patients in the study by presence or absence of wheeze on admitting examination.

	Wheeze on examination			
	Wheeze	**No wheeze**	**All children**	**p**[a]
Total	409 (44%)	529 (56%)	938	
Male sex	266 (65%)	315 (60%)	581 (62%)	0.100
Age				
Age in years (median, IQR)	0.9 (0.5, 1.8)	1.7 (0.7, 3.5)	1.3 (0.6, 2.7)	< 0.001
2 months to <2 years	320 (78%)	291 (55%)		
2 years to <5 years	75 (18%)	140 (26%)		
5 years to 14 years	14 (3.4%)	98 (19%)		
Clinical features[b]				
Fast breathing pneumonia	313 (79%)	290 (67%)	603 (73%)	< 0.001
Chest indrawing pneumonia	112 (28%)	85 (20%)	197 (24%)	0.004
Chest indrawing or fast breathing pneumonia	321 (81%)	301 (70%)	622 (75%)	< 0.001
General danger signs[c]	78 (19%)	148 (28%)	226 (24%)	0.002
Laboratory results				
CRP mg/l (median, IQR)[d]	9.1 (2.6, 22.7)	28 (6.8, 102)	15 (4.1, 52)	< 0.001
Neutrophil count x 10^3/mm^3 (median, IQR)	5.9 (3.8, 9.1)	7.9 (4.5, 14.1)	7.0 (4.0, 11.6)	< 0.001
Radiographic result				
Endpoint consolidation	88 (22%)	238 (45%)	326 (35%)	< 0.001
Infiltrates	64 (16%)	58 (11%)	122 (13%)	0.044
No EPC, uninterpretable for infiltrates	9 (2.2%)	1 (0.2%)	10 (1.1%)	
Normal	236 (58%)	203 (38%)	439 (47%)	
Uninterpretable, not done	12 (3.0%)	23 (4.3%)	35 (3.7%)	
NP carriage of pneumococcal serotypes				
Prior antibiotic therapy	195 (48%)	288 (54%)	483 (51%)	0.047
Carriage of any pneumococci	123 (30%)	197 (37%)	320 (34%)	0.026
Carriage of any pneumococci if no prior antibiotics	94/214 (44%)	127/241 (53%)	221/455 (49%)	0.080
10-valent PCV (VT) pneumococci	32 (7.8%)	91 (17%)	123 (13%)	< 0.001
Additional 13-valent PCV pneumococci	18 (4.4%)	13 (2.5%)	31 (3.3%)	0.142
Non-PCV pneumococci	54 (13%)	66 (12%)	120 (13%)	0.817
Non-typeable pneumococci	19 (4.6%)	27 (5.1%)	46 (4.9%)	0.865
IBD				
Invasive pneumococcal disease[e]	1 (0.2%)	16 (3.0%)	17 (1.8%)	< 0.001
Other invasive pathogens[f]	4 (1%)	8 (1.5%)	12 (1.3%)	0.668

[a]p values calculated using χ^2 test without continuity corrections for two groups as necessary, or using a Welch two-sample t-test of $\log_{10}$ transformed continuous variables (age, CRP and blood neutrophil count).

[b]825/938 cases complete for fast breathing and chest indrawing.

[c]Danger signs: unable to drink, persistent vomiting, convulsions, lethargy or reduced consciousness, stridor in a calm child or severe malnutrition (weight-for-age z-score <-3).

[d]Measured in 474 children

[e]isolated from blood or pleural fluid

[f]In wheeze, no wheeze: *Acinetobacter* spp. 1, 2; *Escherichia coli* 1, 0; *Neisseria meningitidis* 0, 1; *Pseudomonas* spp. 1, 1; *Salmonella enterica* var Typhi 0, 1; *Staphylococcus aureus* 1, 2; viridans streptococci 0, 1.

Table 3.5: Numbers of admission radiographs by radiographic classification and details on the concordance between readers (A) and each reader and the arbiter (B and C) on the diagnosis of endpoint consolidation on admission chest radiographs from children enrolled in the study.

A	**Reader 2**		
Reader 1	**Endpoint consolidation**	**Normal/other infiltrates**	**Total**
Endpoint consolidation	270	69	339
Normal/other infiltrates	58	490	548
Total	328	559	887
Observed agreement:	85.6%		
κ**:**	0.69 (95% CI 0.62, 0.81)		
Test for symmetry:	p = 0.33		
B	**Arbiter**		
Reader 1	**Endpoint consolidation**	**Normal/other infiltrates**	**Total**
Endpoint consolidation	122	46	168
Normal/other infiltrates	33	172	205
Total	155	218	373
Observed agreement:	78.8%		
κ**:**	0.57 (95% CI 0.47, 0.67)		
Test for symmetry:	p = 0.14		
C	**Arbiter**		
Reader 2	**Endpoint consolidation**	**Normal/other infiltrates**	**Total**
Endpoint consolidation	133	28	161
Normal/other infiltrates	25	191	216
Total	158	219	377
Observed agreement:	85.9%		
κ**:**	0.71 (95% CI 0.61, 0.81)		
Test for symmetry:	p = 0.68		

from children with fast breathing pneumonia (98/122, 86%) in comparison to endpoint consolidation (188/326, 37%) or normal radiographs (282/439, 49%, chi^2 test, p < 0.001). As noted previously, endpoint consolidation was associated with absence of wheeze. Endpoint consolidation was associated with higher neutrophil count (median 9.6 x 10^3/mm^3 vs 6.4 x 10^3/mm^3, vs 5.8 x 10^3/mm^3, in endpoint consolidation, infiltrates and normal radiographs, ANOVA, p < 0.001), and higher plasma CRP concentrations (median 51 mg/l vs 16 mg/l, vs 8 mg/l, ANOVA, p < 0.001). See Table 3.5 for details of the concordance of readers for diagnostic endpoints.

Of the 149 children with radiographic endpoint consolidation <2 years of age, 30 (20%) did not have fast breathing or chest indrawing (i.e. did not meet WHO clinical criteria for the diagnosis of pneumonia). Of the 100 children with endpoint consolidation 2 years to 5 years of age, 26 (26%) did not meet WHO clinical criteria for the diagnosis of pneumonia. The specificity of the WHO clinical criteria for endpoint pneumonia could not be calculated:

children who met the WHO clinical critieria for pneumonia, but in whom an alternative diagnosis was made, were not enrolled in the Inpatient Pneumonia Study.

Prior antibiotic therapy or NP carriage of pneumococci was not associated with radiographic endpoint consolidation. However, the proportion of children with NP carriage of VT pneumococci was higher in children with endpoint consolidation (56/326, 17%) vs those with infiltrates (12/122, 9.8%) or normal radiographs (48/439, 11%, chi^2 test, p = 0.021). Invasive pneumococcal disease was identified in 16/326 (4.9%) children with radiographic endpoint consolidation, and in one child with a normal chest radiograph (1/439, 0.2%, p < 0.001). Further characteristics of patients in the study by classification of chest radiograph are in Table 3.6.

Nasopharyngeal carriage of pneumococcal serotypes in pneumonia

Of 938 children enrolled in the study, 320 (34%) had pneumococci cultured from NP swabs. 483/938 (51%) received antibiotic therapy prior to NP sampling (according to parental report) or in-hospital antibiotics. In the 455 children enrolled in the study who had no history of prior antibiotic therapy, 221 (49%) had pneumococci cultured from NP swabs, in comparison to 99/483 (20%) of children who had a history of prior antibiotic treatment (χ^2 test, p < 0.001). NP carriage of vaccine-type pneumococci was associated with older age (median 1.8 years, vs 1.1 years for non-vaccine-type carriage, vs 1.2 years for no carriage, χ^2 test of trend, p = 0.003). Of children <2 years of age, 69/611 (11%) had carriage of vaccine-type pneumococci in comparison to 28/215 (13%) children 2 to 5 years of age and 26/112 (23%) children 5 to 14 years of age.

A lower proportion of children with carriage of vaccine-type pneumococci had fast breathing in comparison with children with carriage of non-vaccine-type pneumococci, or no carriage (61/97, 63% vs 135/186, 73% vs 407/542, 75%, χ^2 test, p = 0.04). A lower proportion of children with carriage of vaccine-type pneumococci had chest indrawing in comparison to children with carriage of non-vaccine-type pneumococci, or no carriage (16/97, 16% vs 34/186, 18% vs 147/542, 27%, χ^2 test, p = 0.010).

Absence of wheeze on examination was also associated with NP carriage of vaccine-type-pneumococci (91/123, 74%) in comparison to carriage of non-vaccine-type pneumococci

Table 3.6: Characteristics of patients enrolled in the study by radiographic classification

	Radiographic result				
	Endpoint consolidation	**Infiltrates**	**Normal**	**All interpretable[a]**	**p[b]**
Total	326 (37%)	122 (14%)	439 (49%)	887	
Male sex	184 (56%)	70 (57%)	292 (67%)	546 (62%)	0.011
Age					
Age in years (median, IQR)	2.3 (1.0, 4.7)	1.1 (0.6, 2.0)	0.9 (0.7, 1.8)	1.3 (0.6, 2.7)	< 0.001
2 months to <2 years	149 (26%)	91 (16%)	338 (58%)	578 (65%)	
2 years to <5 years	100 (49%)	23 (11%)	81 (40%)	204 (23%)	
5 years to 14 years	77 (73%)	8 (8%)	20 (19%)	105 (12%)	
Clinical features[c]					
Fast breathing pneumonia	188 (76%)	98 (86%)	282 (67%)	568 (73%)	< 0.001
Chest indrawing pneumonia	60 (24%)	26 (23%)	99 (24%)	185 (24%)	0.965
Chest indrawing or fast breathing pneumonia	193 (78%)	100 (88%)	294 (70%)	587 (75%)	< 0.001
General danger signs[d]	80 (25%)	29 (24%)	102 (23%)	211 (24%)	0.916
No wheeze on examination	234 (72%)	55 (45%)	202 (46%)	491 (55%)	< 0.001
Laboratory results					
CRP mg/l (median, IQR)[e]	51 (14, 147)	16 (6.0, 38)	8 (2.3, 24)	16 (4.2, 53)	< 0.001
Neutrophil count x 10^3/mm^3 (median, IQR)	9.6 (5.5, 16)	6.4 (3.8, 10)	5.8 (3.8, 9.1)	7.0 (4.0, 12)	< 0.001
NP carriage of pneumococcal serotypes					
Prior antibiotic therapy	184 (56%)	63 (51%)	214 (49%)	461 (52%)	0.108
Carriage of any pneumococci	117 (36%)	41 (34%)	144 (33%)	302 (34%)	0.668
Carriage of any pneumococci if no prior antibiotics	77/139 (55%)	29/58 (50%)	98/223 (44%)	204/420 (49%)	0.103
10-valent PCV (VT) pneumococci	56 (17%)	12 (9.8%)	48 (11%)	116 (13%)	0.021
Additional 13-valent PCV pneumococci	9 (2.8%)	6 (4.9%)	13 (3.0%)	28 (3.2%)	0.482
Non-PCV pneumococci	35 (11%)	15 (12%)	65 (15%)	115 (13%)	0.246
Non-typeable pneumococci	17 (5.2%)	8 (6.6%)	18 (4.1%)	43 (4.8%)	0.500
IBD					
Invasive pneumococcal disease[f]	16 (4.9%)	0	1 (0.2%)	17 (1.9%)	<0.001
Other invasive pathogens[g]	4 (1.2%)	2 (1.6%)	3 (0.7%)	9 (1.0%)	0.577

[a]Excludes patients with radiographs not done, or uninterpretable (for infiltrates, or for any feature).

[b]p values calculated using χ^2 test without continuity corrections for two or three groups as necessary, or using one-way analysis of variance (ANOVA) of $\log_{10}$ transformed continuous variables (age, CRP and blood neutrophil count).

[c]825/938 cases complete for fast breathing and chest indrawing.

[d]Danger signs: unable to drink, persistent vomiting, convulsions, lethargy or reduced consciousness, stridor in a calm child or severe malnutrition (weight-for-age z-score <-3).

[e]Measured in 474 children

[f]isolated from blood or pleural fluid

[g]In radiographic groups: endpoint consolidation, infiltrates, normal: *Acinetobacter* spp. 1, 1, 0; *Escherichia coli* 0, 0, 1; *Neisseria meningitidis* 0, 0, 1; *Pseudomonas* spp. 1, 0, 0; *Salmonella enterica* var Typhi 0, 0, 1; *Staphylococcus aureus* 2, 1, 0; viridans streptococci 0, 0, 0.

(106/197, 54%) or no carriage (332/618, 54%, chi^2 test, p < 0.001). Carriage of vaccine-type pneumococci was associated with higher neutrophil count (median 10.4 x 10^3/mm^3) in comparison to carriage of non-vaccine-type pneumococi (7.7 x 10^3/mm^3) or no carriage (6.4 x 10^3/mm^3, ANOVA, p < 0.001), and higher plasma CRP concentrations (median 51 mg/l) in comparison to carriage of non-vaccine-type pneumococci (16 mg/l) or no carriage (8 mg/l, ANOVA, p < 0.001).

The association of absence of wheeze on examination with NP carriage of vaccine-type pneumococci was greatest for NP carriage of serotypes 1 (17 children without wheeze, 2 children with wheeze) and 5 (10 children without wheeze, 0 children with wheeze). Of pneumococcal serotypes carried in $\geq$5 children, serotypes 5 and 1 were associated with the greatest blood neutrophil count (median 18 x 10^3/mm^3 and 14 x 10^3/mm^3, respectively). These data are detailed in Table 3.8

Invasive disease potential of pneumococcal serotypes

The proportional NP carriage of pneumococcal serotypes in all enrolled children admitted with pneumonia is described, by age group, in Figure 3.7. Excluding non-typeable pneumococci, serotype 14 had the greatest carriage prevalence in children <5 years of age. Serotype 1 had the greatest carriage prevalence in children 5 to 14 years of age.

For comparison, the proportional NP carriage of pneumococcal serotypes in children in the community, enrolled into a separate but related study (the Community Carriage Study: children 6 months to <2 years of age, n = 2895; children 2 to 5 years of age, n = 307, see Section 2.3.4) is described in Figure 3.8. Excluding non-typeable pneumococci, serotype 19F had the greatest carriage prevalence in children 6 months to <2 years of age. Serotype 14 had the greatest carriage prevalence in children 2 to 5 years of age. Serotype 1 was rarely carried in either age group. (Pending serotypes are rare serotypes which require further serotyping reagents, and are all typeable but not contained within PCVs.)

The odds ratio for NP carriage of pneumococcal serotypes contained within 13-valent PCV (i.e. 10-valent PCV plus serotypes 3, 6A and 19A) and non-typeable pneumococci

Table 3.7: Characteristics of patients enrolled in the study by carriage of vaccine-type (10-valent PCV), non-vaccine-type or no NP carriage of pneumococci.

	NP carriage of pneumococci				
	Vaccine-type	**Non-vaccine-type**	**None**	**All children**	**p[a]**
Total	123 (13%)	197 (21%)	618 (66%)	938	
Male sex	75 (61%)	106 (54%)	400 (65%)	581 (62%)	0.022
Age					
Age in years (median, IQR)	1.8 (0.7, 2.2)	1.1 (0.5, 2.2)	1.2 (0.6, 2.7)	1.3 (0.6, 2.7)	0.003
2 months to <2 years	69 (11%	141 (23%)	401 (66%)	611 (65%)	
2 years to <5 years	28 (13%)	46 (21%)	141 (66%)	215 (23%)	
5 years to 14 years	26 (23%)	10 (8.9%)	76 (68%)	112 (12%)	
Clinical features[b]					
Fast breathing pneumonia	61 (63%)	135 (73%)	407 (75%)	603 (75%)	0.044
Chest indrawing pneumonia	16 (16%)	34 (18%)	147 (27%)	197 (21%)	0.010
Chest indrawing or fast breathing pneumonia	65 (67%)	137 (73%)	420 (77%)	622 (77%)	0.072
General danger signs[c]	30 (24%)	46 (23%)	150 (24%)	226 (24%)	0.963
No wheeze on examination	32 (73%)	91 (52%)	286 (53%)	409 (44%)	< 0.001
Laboratory results					
CRP mg/l (median, IQR)	32 (8.2, 148)	21 (9.6, 38)	12.8 (2.9, 41)	16 (4.1, 52)	< 0.001
Neutrophil count x 10^3/mm^3 (median, IQR)	10.4 (5.2, 17)	7.7 (4.4, 12)	6.4 (3.9, 10)	7.0 (4.0, 12)	< 0.001
Radiographic results					
Endpoint consolidation	56 (46%)	61 (31%)	209 (34%)	326 (35%)	<0.020
Infiltrates only	12 (9.8%)	29 (15%)	81 (13%)	122 (13%)	0.435
No EPC, uninterpretable for infiltrates	0	0	10 (1.6%)	10 (1.1%)	
Normal	338 (55%)	81 (38%)	20 (18%)	439 (47%)	
Uninterpretable, not done	7 (5.7%)	11 (5.6%)	20 (3.2%)	35 (3.7%)	
Prior antibiotic therapy	39 (32%)	60 (30%)	384 (62%)	483 (51%)	< 0.001
IBD					
Invasive pneumococcal disease[d]	6 (4.9%)	2 (1.0%)	9 (1.5%)	17 (1.9%)	<0.022
Other invasive pathogens[e]	0	3 (1.5%)	9 (1.5%)	12 (1.3%)	0.400

[a] p values calculated using χ^2 test without continuity corrections for two or three groups as necessary, or using one-way analysis of variance (ANOVA) of $\log_{10}$ transformed continuous variables (age, CRP and neutrophils count).

[b] 825/938 cases complete for fast breathing and chest indrawing.

[c] Danger signs: unable to drink, persistent vomiting, convulsions, lethargy or reduced consciousness, stridor in a calm child or severe malnutrition (weight-for-age z-score <-3).

[d] isolated from blood or pleural fluid

[e] In radiographic groups: endpoint consolidation, infiltrates, normal: *Acinetobacter* spp. 0, 0, 3; *Escherichia coli* 0, 0, 1; *Neisseria meningitidis* 0, 1, 0; *Pseudomonas* spp. 0, 0, 2; *Salmonella enterica* var Typhi 0, 1, 0; *Staphylococcus aureus* 0, 0, 3; viridans streptococci 0, 1, 0.

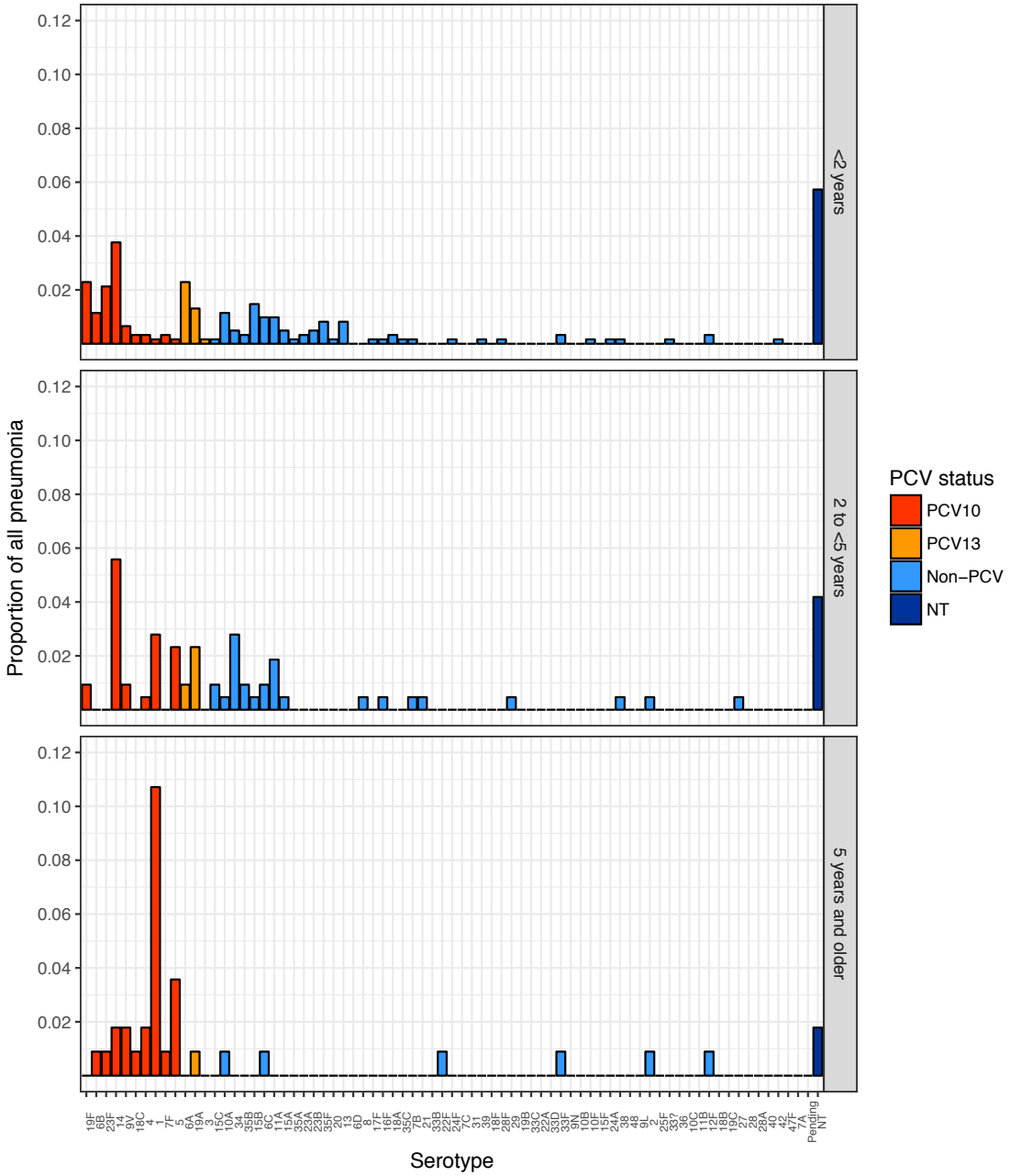


Figure 3.7: Nasopharyngeal carriage of serotypes of pneumococci in enrolled children <2 years of age (n = 611), 2 to 5 years of age (n = 215) and 5 to 14 years of age (n = 112) admitted to Patan Hospital with a clinician diagnosis of pneumonia, March 2014–December 2016. Serotypes are ordered according to their prevalence in children in the community (see Figure 3.8) and by their presence in 10-valent PCV (PCV10) and 13-valent PCV (PCV13).

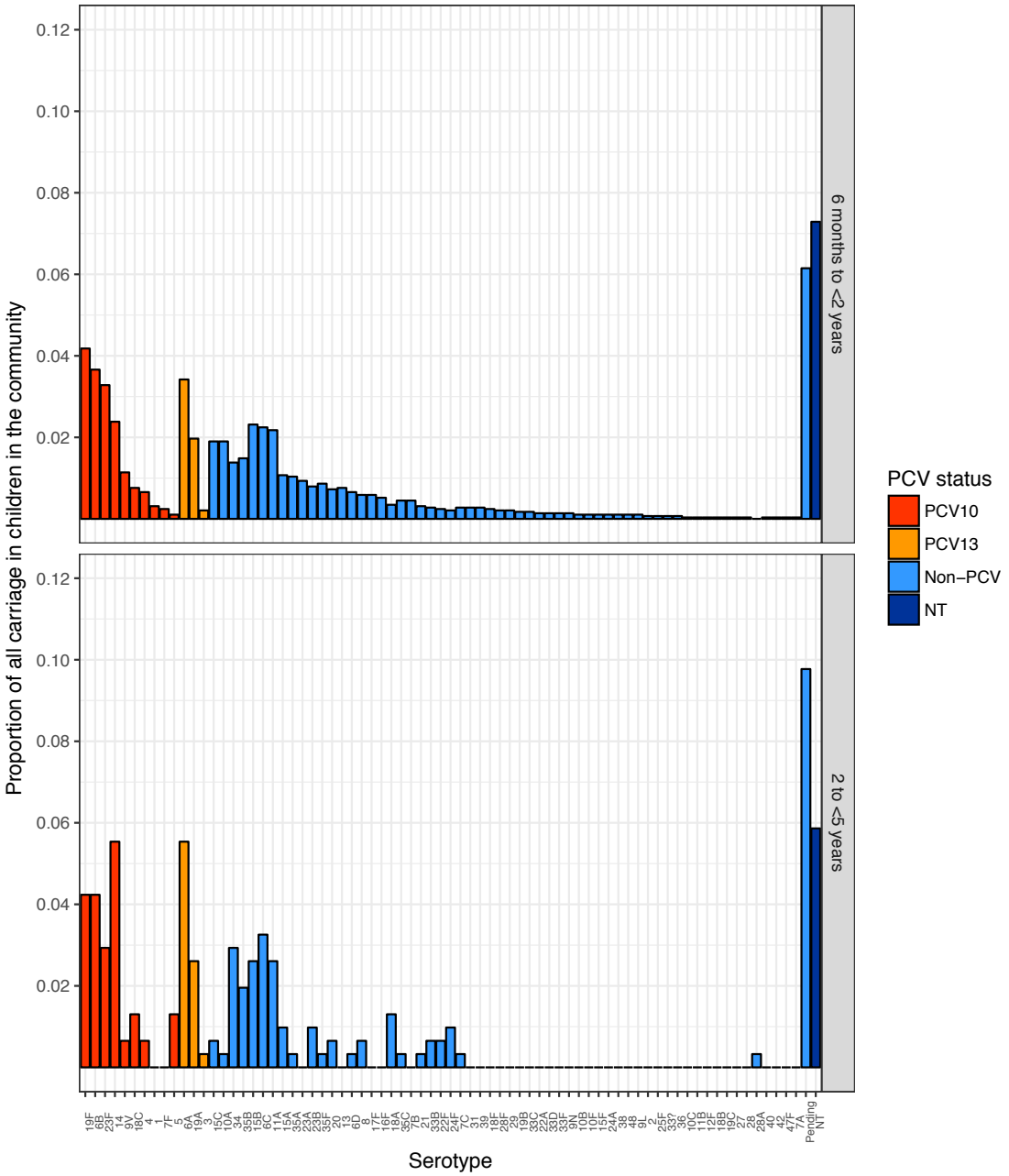


Figure 3.8: Nasopharyngeal carriage of serotypes of pneumococci in children 6 months to 2 years of age (n = 2895) and 2 to 5 years of age (n = 307) in the community, according to age group, March 2014–December 2016. Serotype are ordered by proportion and their presence in 10-valent PCV (PCV10) and 13-valent PCV (PCV13).

Table 3.8: Association of NP carriage of pneumococcal serotypes within 10-valent and 13-valent PCV with absence of wheeze, and blood neutrophil count in children enrolled to the Inpatient Pneumonia Study at Patan Hospital.

Serotype	PCV	NP carriage and absence of wheeze	NP carriage and wheeze	Odds ratio (unadjusted)	Neutrophil count x 10^3/mm^3 (median, IQR)
1	10-valent	17	2	8.5	14 (10, 19)
4	10-valent	3	2	1.5	14 (2.6, 27)
5	10-valent	10	0	∞	18 (7.0, 25)
6A	10-valent	5	11	0.45	10 (5.1, 15)
7F	10-valent	2	1	2	17 (9.0, 20)
9V	10-valent	5	3	1.7	8.9 (6.5, 12)
14	10-valent	26	11	2.4	11 (7.6, 17)
18C	10-valent	3	0	∞	5.1 (4.4, 5.5)
19F	10-valent	10	6	1.7	12 (4.6, 14)
23F	10-valent	8	6	1.3	7.1 (3.7, 8.4)
3	13-valent	0	1	0	1.3
6B	13-valent	7	1	7	5.3 (3.5, 9.4)
19A	13-valent	8	6	1.3	7.4 (4.0, 14)
NT	None	27	19	1.4	8.6 (4.4, 10.4)

Table 3.9: Odds ratios for NP carriage of pneumococcal serotypes contained within 13-valent PCV and non-typeable (NT) pneumococci in pneumonia and community control children, adjusted for age and sex, *prior* to the introduction of 10-valent PCV into routine infant immunisation schedule of the Kathmandu Valley.

Serotype	Adjusted odds ratio (95% CI)
1	5.0 (1.7, 14.0)
3	1.0 (0.05, 5.9)
4	1.5 (0.32, 5.1)
5	4.0 (0.31, 37.6)
6A	0.45 (0.17, 0.96)
6B	0.19 (0.04, 0.54)
7F	1.1 (0.04, 8.7)
9V	1.0 (0.28, 2.9)
14	1.5 (0.81, 2.8)
18C	0.36 (0.02, 1.8)
19A	0.68 (0.20, 1.7)
19F	0.33 (0.11, 0.73)
23F	0.54 (0.22, 1.1)
NT	1.2 (0.75, 1.9)

in children with pneumonia vs children in the community are described in Table 3.9. Carriage of serotype 1 was significantly associated with pneumonia when adjusted for age and sex (odds ratio 5.0, 95% CI 1.7, 14.0). No other serotype was significantly associated with pneumonia (Figure 3.9). NP carriage of serotype 1 pneumococci was highly associated with increasing age: 1/611 (0.2%) children 2 months to <2 years of age, 6/210 (1.6%) children 2 years to <5 years of age and 12/99 (12%) children $\geq$5 years of age.

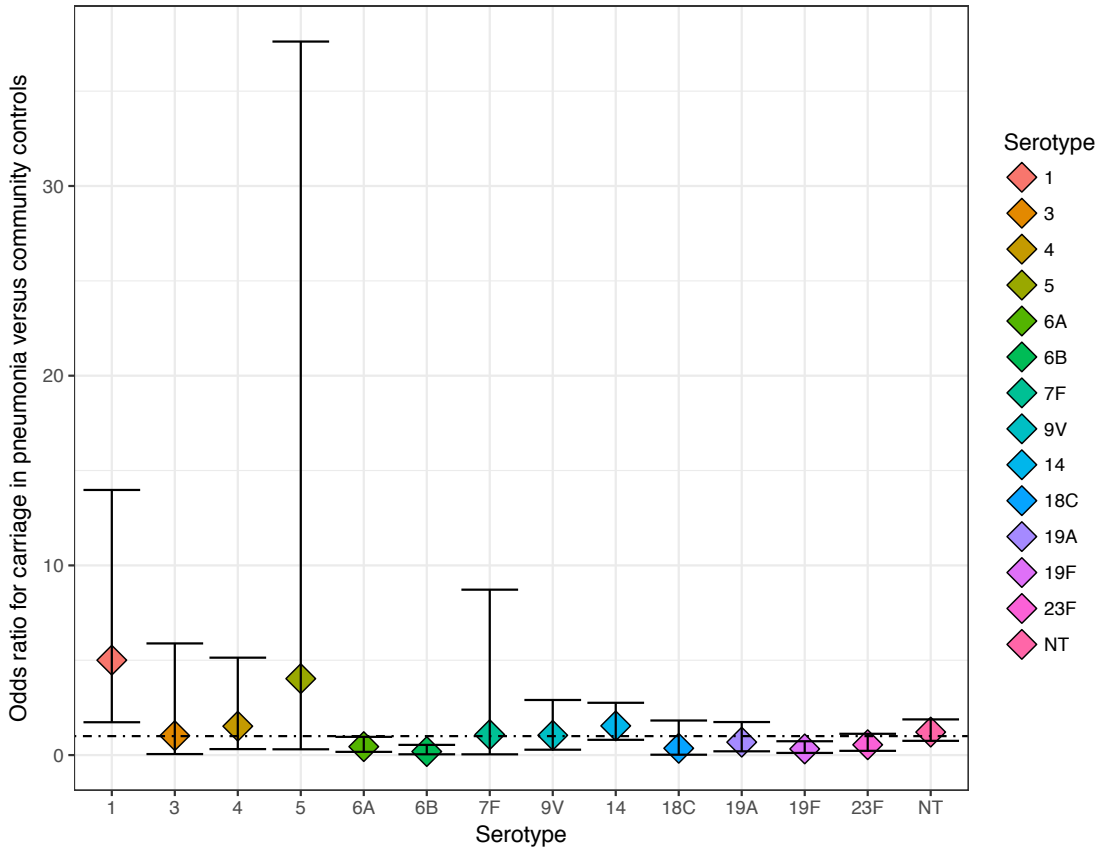


Figure 3.9: Odds ratios for NP carriage of pneumococcal serotypes contained within PCV13 and non-typeable (NT) pneumococci in pneumonia and community control children (all ages), adjusted for age and sex, prior to introduction of 10-valent PCV into the Kathmandu Valley. Error bars show 95% CIs.

Predictors of invasive pneumococcal disease derived from children with pneumonia

Variables that were statistically associated, biologically plausible, and measured in all age groups of children with invasive pneumococcal disease (absence of wheeze on examination, radiographic endpoint consolidation, plasma CRP concentration, blood neutrophil count and age of child) or associated with pneumonia in comparison to community controls (NP carriage of pneumococcal serotype 1) are shown in Table 3.10. In logistic regression analyses adjusted for age, radiographic endpoint consolidation (vs non endpoint consolidation) had the highest odds ratio (OR) for invasive pneumococcal disease (OR 19, 95% CI 2.5, 153), followed by NP carriage of pneumococcal serotype 1 (OR 5.3, 95% CI 1.2, 23) and CRP concentration (OR 1.1, 95% CI 1.1, 1.2, for each 10 mg/l increase).

The sensitivity and specificity of endpoint consolidation for invasive pneumococcal

Table 3.10: Odds ratios (OR) for association between clinical features, laboratory results, and NP carriage of pneumococcal serotypes, unadjusted and adjusted for age of child for invasive pneumococcal disease (IPD).

	Unadjusted OR for IPD	**OR for IPD (adjusted for age)**
Absence of wheeze on examination	6.2 (1.4, 27, p = 0.016)	3.5 (0.8, 16, p = 0.109)
Radiographic endpoint consolidation	32 (4.2, 239, p < 0.001)	19 (2.5, 153, p = 0.005)
CRP concentration/10 mg/l increase	1.2 (1.1, 1.2, p < 0.001)	1.1 (1.1, 1.2, p < 0.001)
Blood neutrophil count/10^3/mm^3 increase	1.1 (1.0, 1.1, p < 0.001)	1.0 (1.0, 1.1, p = 0.065)
NP carriage of serotype 1	13 (3.1, 51, p < 0.001)	5.3 (1.2, 23, p = 0.03)
NP carriage of other VT serotypes	2.0 (0.5, 7.6, p = 0.301)	2.0 (0.5, 7.9, p = 0.310)
Age (per increase of 1 year)	1.3 (1.2, 1.4, p < 0.001)	–

disease was 0.94 (95% CI 0.71, 1.00) and 0.66 (95% CI 0.63, 0.69), respectively. The positive predictive value (PPV) of endpoint consolidation was 0.05 (95% CI 0.4, 0.68), and the negative predictive value (NPV) 0.99 (95% CI 0.99, 1.00). The sensitivity of NP carriage of pneumococcal serotype 1 for invasive pneumococcal disease was 0.17 (95% CI 0.04, 0.43) and specificity of 0.98 (95% CI 0.97, 0.99), with a PPV of 0.16 (95% CI 0.10, 0.40) and a NPV of 0.98 (95% CI 0.92, 0.99). Analysis of receiver operating characteristics (ROC) curves (using Youden's index) for CRP concentration produced an optimal threshold of 58 mg/l, with a sensitivity of 0.90 (95% CI 0.55, 1.00), specificity of 0.78 (95% CI 0.74, 0.82), PPV of 0.08 (95% CI 0.07, 0.80), and NPV of 0.99 (95% CI 0.98, 1.00).

These variables were entered into classification and regression tree models for the prediction of invasive pneumococcal disease in children in the study, using the R package $partykit$[232]. This technique has been pioneered in data mining and extended into the assessment of diagnostic tests for the aetiology of infection[233].

The classification tree shown in Figure 3.10 uses the variables absence of wheeze on examination, blood neutrophil count, radiographic endpoint consolidation and NP carriage of pneumococcal serotype 1. These variables were measured on almost all children across all years of the study. In these data, the presence of radiographic endpoint consolidation was the best predictor of invasive pneumococcal disease (in 16/326, 4.9%) children with endpoint consolidation and 1/612 (0.2% of other children).

The classification tree following the addition of plasma CRP concentration (measured on 474 children since February 2015) is shown in Figure 3.11. In this model, a

CRP concentration >249.19 mg/l, when combined with NP carriage of serotype 1 pneumococci classified invasive pneumococcal disease accurately (node 9, n = 2). A CRP concentration of >249.19 without carriage of serotype 1 classified invasive pneumococcal disease in two further children (node 8, 13.3% of n = 15). Children with lower CRP concentrations but with a neutrophil count of >19.36 x 10^3/mm^3 accounted for a further four cases of invasive pneumococcal disease (node 7, 5.5% of n = 73). Children with blood neutrophil count $\leq$19.36 and CRP concentration >180.35 accounted for two further cases (node 5, 4.7%, n = 43). The lowest proportion of children with invasive pneumococcal disease was in the group of children with CRP concentrations $\leq$180.35 (node 4, 0.9%, n = 805). However, node 4 contained seven children with invasive pneumococcal disease among the large majority of all children enrolled to the study (805/938, 86%).

Addition of age to the classification tree model is shown in Figure 3.12. In this model, a CRP concentration >249.19 and age >7.513 years classified invasive pneumococcal disease in two children (node 9, 40%, n = 5). A CRP concentration >249.19 and age $\leq$7.513 classified invasive pneumococcal disease in two further children (node 8, 17%, n = 15). The classification for children with a CRP concentration >249.19 is identical between Figure 3.12 and Figure 3.11. In brief, the addition of age to the conditional inference model in Figure 3.12 replaces the decision node that uses data on presence or absence of carriage of serotype 1 pneumococci (Figure 3.11).

3.4 Discussion

3.4.1 Invasive pneumococcal disease

Pneumococci were the second-most prevalent bacteria isolated from normally sterile sites in children admitted to Patan Hospital with suspected IBD (pneumonia, sepsis, and meningitis) from March 2005 to December 2016.

During the period 2010–2016 (when data were collected on all children with suspected IBD) bacterial pathogens were isolated from the blood of 3.3% of those enrolled, with invasive pneumococcal disease present in 0.6% of those enrolled (17% of all "positive"

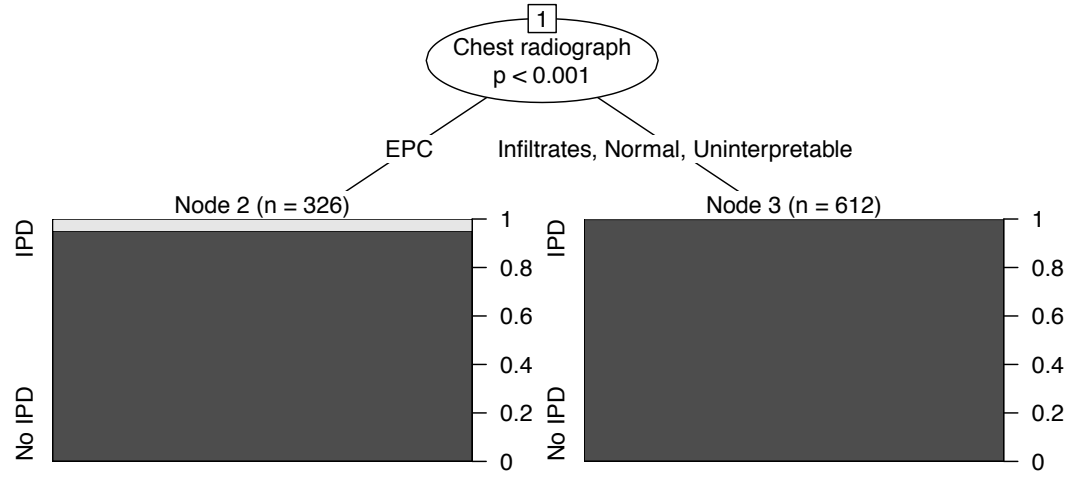


Figure 3.10: Conditional inference tree for the prediction of invasive pneumococcal disease in children with pneumonia at Patan Hospital using the variables: **absence of wheeze** on examination, **blood neutrophil count** x 10^3/mm^3, **chest radiograph results**, and **NP carriage of pneumococcal serotypes** (as serotype 1, other VT, non-VT and none). Missing data were not included, and each node required p < 0.05 in order for a split to occur.

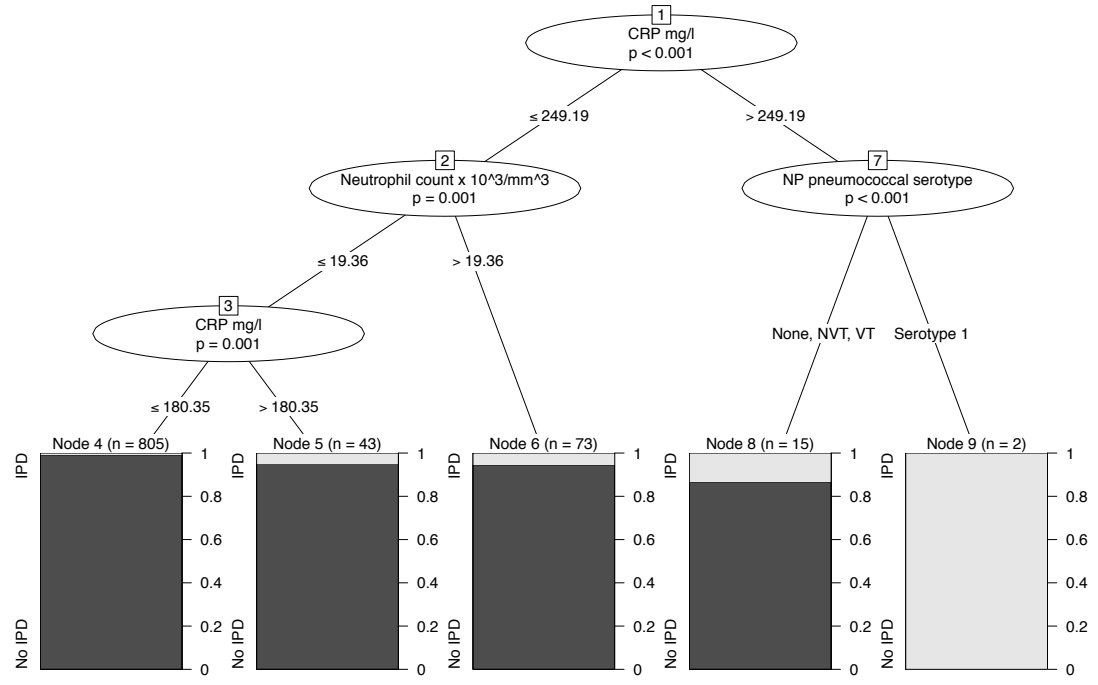


Figure 3.11: Conditional inference tree for the prediction of invasive pneumococcal disease in children with pneumonia at Patan Hospital using the variables: **absence of wheeze on examination**, **CRP (mg/l)**, **blood neutrophil count** x 10^3/mm^3, **chest radiograph results**, and **NP carriage of pneumococcal serotypes** (as serotype 1, other VT, non-VT and none). Missing data were not included, and each node required p < 0.05 in order for a split to occur.

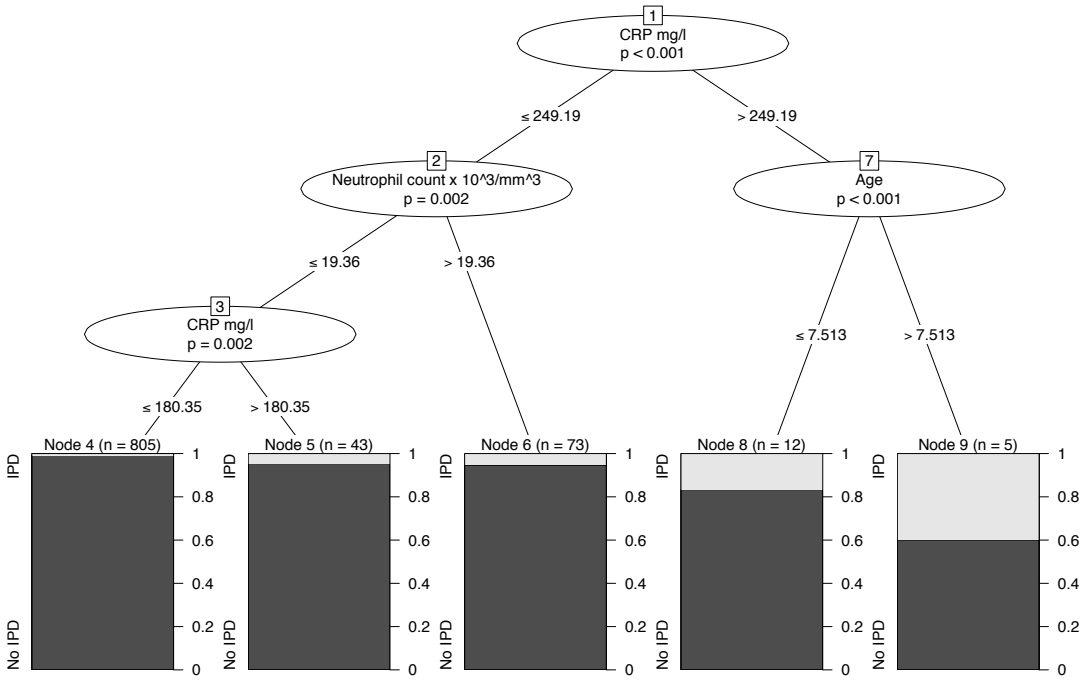


Figure 3.12: Conditional inference tree for the prediction of invasive pneumococcal disease in children with pneumonia at Patan Hospital using the variables: **age of child**, **absence of wheeze** on examination, **CRP (mg/l)**, **blood neutrophil count** x 10^3/mm^3, **chest radiograph results**, and **NP carriage of pneumococcal serotypes** (as serotype 1, other VT, non-VT and none). Missing data were not included, and each node required p < 0.05 in order for a split to occur. See Figure 3.11 for comparison without age in the tree.

blood cultures). The majority of invasive pneumococcal disease was identified in children $\geq$2 years of age, during this period. Only half of isolates of pneumococci from children <2 years of age were of a serotype contained within 10-valent PCV, with this proportion rising with age, such that nearly 9/10 isolates from children $\geq$5 years of age were contained within the vaccine. Prior antibiotic therapy in $\sim$50% of children, across all age groups, is likely to have considerably lowered the prevalence of pathogens isolated from blood in this study.

The prevalence of pneumococci as the cause of childhood IBD in these data is slightly higher than that of other studies in South and Southeast Asia, where approximately 8–13% of positive blood cultures in children are caused by *S. pneumoniae*[77, 234, 235]. Varying clinical criteria for "suspected" IBD across different sites may account for some of these differences, as will socioeconomic factors, and the presence of other

important causes of IBD (notably typhoidal Salmonellae in this case series). However, the prevalence of invasive pneumococcal disease described here is considerably lower than that described in data from sub-Saharan Africa, where approximately one quarter of childhood IBD is caused by *S. pneumoniae*[78, 81], and is associated with higher prevalence of HIV in children in these settings[235, 236]. Data were not routinely collected on the presence or absence of malnutrition (or of weight on admission) in this case series of children with suspected IBD. However, one might hypothesise that an increased prevalence of severe malnutrition (in both children with and without HIV infection) might account for significant differences in prevalence of invasive pneumococcal disease in children admitted to hospital with pneumonia in Africa.

Given the low sensitivity of blood culture for pneumococci (and other pathogens), pneumococcal disease is likely to be a major cause of disease in children admitted to Patan Hospital with suspected IBD. Nonetheless, proven invasive pneumococcal disease is epidemiologically rare, making it a poor endpoint for assessment of the impact of 10-valent PCV on pneumococcal disease. Children $\geq$2 years of age have a considerably higher prevalence of invasive pneumococcal disease than younger children at Patan Hospital. This compounds the difficulty of using invasive pneumococcal disease as an endpoint for vaccine impact in these data. In the absence of indirect effects of 10-valent PCV the current birth cohort of children vaccinated with 10-valent PCV would need to be 2 years to 5 years of age before the impact of PCV on invasive pneumococcal disease is detectable. Even with indirect vaccine effects (for example, by limiting transmission of vaccine-type pneumococci from vaccinated infants to unvaccinated older children) one would hypothesise that a period of high vaccine uptake would be needed to detect the impact of 10-valent PCV on invasive pneumococcal disease prevalence in the childhood IBD surveillance programme.

3.4.2 The Inpatient Pneumonia Study

The Inpatient Pneumonia Study provides detailed clinical, radiographic and microbiological data on children admitted to Patan Hospital with pneumonia from March 2014

to December 2016. Enrolment is currently continuing (as of August 2017). Refusal of consent for enrolment to the study was very low, and these data thus represent a systematic case series of pneumonia in children admitted to Patan Hospital during the study period.

The clinical diagnosis of pneumonia for children enrolled to the Inpatient Pneumonia Study was based upon *IMCI* guidelines[94], and the WHO clinical criteria for the diagnosis of pneumonia in healthcare facilities[15], in children <5 years of age, but with a high degree of clinical discretion by the admitting paediatric team. In children aged 5 years to 14 years, the diagnosis of pneumonia was based upon clinical discretion by the admitting paediatric team only. To limit varying thresholds for the diagnosis, and admission, of children with pneumonia all admissions were reviewed at a daily meeting at which all admitting paediatricians (including consultants) were in attendance.

In this case series, the sensitivity of WHO clinical criteria for pneumonia (cough or difficulty breathing, with fast breathing and/or chest indrawing) was low: 20% of children <2 years of age, and 26% of children aged 2 years to <5 years of age with radiographic endpoint consolidation would have been missed using these criteria for the diagnosis of pneumonia. This implies that, in this setting and population of children at a tertiary teaching hospital in Kathmandu, either WHO clinical criteria for the diagnosis of pneumonia, or the clinical identification of signs related to the criteria, lacked sensitivity. Although all nurses and doctors were trained in WHO IMCI guidelines, I cannot exclude thepossibility that some clinical signs may have been overlooked during this decade-long case series.The inclusion of children 5 to 14 years of age (in comparison to a number of other studies of childhood pneumonia, including the PERCH study[10]), identified a large proportion of older children with radiographic endpoint consolidation (69%). In addition, 65% of invasive pneumococcal disease was identified in children within this older age group in this case series.

The generalisability of these data to other settings is limited by the absence of generalisable inclusion criteria: enrolment was, to a significant extent, based upon the discretion of the admitting paediatric team.This also led to the exclusion of children who

did not have a clinical diagnosis of pneumonia (but who otherwise met the WHO clinical criteria). However, the use of clinical discretion in the diagnosis of pneumonia, and the inclusion of children aged 5 to 14 years, identified a significant proportion of children with endpoint consolidation and invasive pneumococcal disease who would otherwise have been missed.

3.4.3 Proxy markers of pneumococcal pneumonia

In analyses adjusted for age, radiographic endpoint consolidation, increased CRP concentration and NP carriage of pneumococcal serotype 1 were all associated with increased odds for invasive pneumococcal disease in this case series.

Radiographic endpoint consolidation was present in 35% of children enrolled to the Inpatient Pneumonia Study. This was lower than previously described in data from the initial study of childhood pneumonia aetiology at Patan Hospital during the period March 2005–December 2006, where 44% of radiographs were deemed to have endpoint consolidation[4]. This decrease may be related to secular trends in pneumonia aetiology or severity, or may represent the impact of vaccination against Hib, introduced in 2009, or secondary to (undescribed) changes in the healthcare-seeking behaviour, or the behaviour of admitting paediatricians during this period. This was considerably higher than that described in the PERCH study, where 27% of children had endpoint consolidation[237]. This may be related to the enrolment of an older age group (5 to 14 years) in the Inpatient Pneumonia Study in comparison to the PERCH study (where the maximum age of enrolment was <5 years). Alternatively, enrolled cases may have had a different severity, or aetiology to children in the PERCH study.

Presence of endpoint consolidation (vs absence of endpoint consolidation) had a sensitivity of 0.94 and a specificity of 0.66 for invasive pneumococcal disease, but with a PPV of 0.04 (i.e. only 4% of children with endpoint consolidation had invasive pneumococcal disease). Data from the initial study of childhood pneumonia aetiology at Patan Hospital described invasive pneumococcal disease in 10/282 (3.5%) children with endpoint consolidation, suggesting that the proportion of childhood pneumonia attributable to

pneumococcal infection may have remained relatively stable during this period prior to and during the recent introduction of 10-valent PCV to the infant immunisation schedule[4]. Data from the PERCH study describing the relationship between radiographic data and the aetiology of childhood pneumonia have not yet been published[237].

An optimum CRP concentration threshold of 58 mg/l for the identification of invasive pneumococcal disease in children with pneumonia is somewhat similar to that derived from the PERCH study to discriminate between any bacterial pneumonia and pneumonia secondary to respiratory syncytial virus infection. In the PERCH study, the optimum threshold was 37 mg/l, with a sensitivity of 0.77 and a specificity of 0.82. Relatively slight differences in this value may reflect these different comparator groups (in particular invasive pneumococcal disease is may typically have a higher CRP than other "bacterial" pneumonias). However, CRP concentration was only measured since February 2015 in the case series, limiting its utility to assess pneumococcal pneumonia over time (see Chapter 4). These data do not allow assessment of any association of CRP concentration with clinical severity due to the absence of any inpatient mortality in the case series, and limited data on other markers of pneumonia severity.

A major goal of this study was the assessment of NP carriage of pneumococcal serotypes in children with pneumonia (and in comparison to age-stratified community controls from the Community Carriage Study, Section 2.3.4). All children enrolled within the study had a NP specimen cultured for pneumococci, with a resulting carriage prevalence of 34%, or 49% in children with no history of antibiotics prior to sampling.

NP carriage of pneumococci, as a whole, was not associated with invasive pneumococcal disease, with endpoint consolidation (Table 3.6) or absence of wheeze on examination (Table 3.4; after exclusion of patients with prior antibiotic treatment). However, NP carriage of vaccine-type pneumococci was associated with absence of wheeze on examination, radiographic endpoint consolidation and invasive pneumococcal disease (Table 3.7), suggesting that these serotypes may be causative in a proportion of these patients. However, this association was largely reliant upon the effect of NP carriage of serotype 1 pneumococci. These data support the hypothesis that NP carriage of vaccine-type

pneumococci in pneumonia, and community controls, could be used as one of a number of endpoints against which to assess the impact of PCV on pneumonia, rather than simply NP carriage *per se*.

To assess the invasive disease potential of individual pneumococcal serotypes, I restricted the analysis to data prior to the introduction of 10-valent PCV. Of vaccine-type pneumococci (10-valent PCV), and the additional 3 serotypes in 13-valent PCV, only NP carriage of pneumococcal serotype 1 was highly associated with invasive pneumococcal disease (OR 5.0), when adjusted for age. Serotype 5 had high odds for carriage in pneumonia vs community controls but the association was non-significant. These results are comparable to data from Israel[67], where serotypes 1 and 5 (and additionally 7F, 22F, 14, 19A and 9V were associated with pneumonia vs healthy controls (unadjusted for age) a previous analysis of isolates from the UK[66] and a global meta-analysis prior to introduction of PCV[65]. NP carriage of serotype 1 pneumococci was highly specific for invasive pneumococcal disease (specificity 0.98), but its low prevalence made limited its utility as a diagnostic test.

NP carriage of serotype 1 pneumococci was highly associated with increased age in the Inpatient Pneumonia Study. In addition, serotype 1 pneumococci were responsible for nearly 50% of invasive pneumococcal disease in the childhood IBD surveillance programme. The importance of serotype 1 pneumococci for invasive pneumococcal disease in South Asia has been recently shown in a meta-analysis of data from children <5 years of age in India[238], where it was the second most prevalent invasive isolate (11%) after serotype 14 (13%). However, there is an important – and under-recognised – association of NP carriage of serotype 1 pneumococci in children $\geq$5 years of age with pneumonia, and invasive pneumococcal disease in our data. A major finding is therefore that identification of childhood invasive pneumococcal disease – and particularly that caused by serotype 1 pneumococci – requires sampling of children $\geq$5 years of age.

The formal incorporation of data regarding plasma CRP concentration, NP carriage of serotype 1 pneumococci, and blood neutrophil count into a classification and regression

tree analysis, resulted in a series of decision nodes that could be used to predict the prevalence of invasive pneumococcal disease (i.e. "definite pneumococcal pneumonia") in different classifications of enrolled children. This analysis forms the basis of the pneumonia aetiology comparator standards used extensively in Chapter 5 and Chapter 6 to assess novel diagnostic tests for the aetiology of pneumonia. With appropriate adjustments (for example, to increase the sensitivity of CRP concentration for bacterial pneumonia by lowering the threshold of "positivity") could also be used to create a treatment algorithm for children at Patan Hospital, potentially reducing unnecessary antibiotic administration and thereby limiting the emergence of antimicrobial resistance. Classification and regression tree analysis also formalises the context in which the additive benefit of a novel diagnostic test can be assessed, given presumed cost constraints[233].

3.4.4 Strengths, limitations, future work and summary

These data are derived from a childhood IBD surveillance programme that has enrolled children with IBD since March 2005. The inclusion of detailed clinical data on all children with suspected IBD from 2010 onwards provides a major resource for the analysis of childhood IBD in South Asia over time. The planned addition of demographic data, and basic clinical diagnoses from all children admitted to Patan Hospital from 2010 onwards (data not shown) will enrich this dataset further.

The low prevalence of invasive pneumococcal disease, and the preponderance of invasive pneumococcal disease in older children (in these data), make this an insensitive endpoint for assessing the impact of 10-valent PCV in Nepal. Nonetheless, one may hypothesise that the prevalence of invasive pneumococcal disease in children with suspected IBD will decrease once the birth cohort of children vaccinated with 10-valent PCV reach the age at which invasive pneumococcal disease is typically detected in children in this case series. This "lag" may be shortened by indirect effects of 10-valent PCV on reducing transmission of vaccine-type pneumococci in the community. Thus, the inclusion of children $\geq$5 years

of age in both the IBD surveillance programme and the Inpatient Pneumonia Study is a major strength of these data.

In this study, approximately one third of enrolled children <5 years of age did not apparently meet the WHO clinical criteria for the diagnosis of pneumonia. Of these children, 29% had radiographic endpoint consolidation (in comparison to 33% of children who did meet the criteria). As discussed above, without clear audit of clinical assessment techniques, if is difficult to confidently attribute low sensitivity to poor performance of the WHO clinical criteria. However, if true, this has important clinical implications – given that the WHO clinical criteria are expected to be highly sensitive for the identification of likely bacterial pneumonia (for which endpoint consolidation has been considered a proxy marker) to guide antibiotic therapy. It is possible to subset these data by children who only met the WHO clinical criteria for pneumonia. The exclusion of children who met the WHO clinical criteria for pneumonia but who were allocated an alternative diagnosis provides a greater challenge to the comparability of these findings to other studies, such as the PERCH study[10].

Both the childhood IBD surveillance programme, and the Inpatient Pneumonia Study provide longitudinal data. These data are explored in Chapter 4. The extension of both of these studies into the future is of critical importance in the assessment of the impact of PCV in Nepal, for reasons discussed above. In addition, detailed study of the NP carriage of pneumococcal serotypes in pneumonia, in community controls, and of invasive pneumococcal disease may provide important data on "serotype-replacement" that has been shown to follow introduction of 7-valent PCV in other settings[239].

4

Impact of pneumococcal conjugate vaccination in Nepal: preliminary assessment

4.1 Introduction

As described earlier, 10-valent PCV was introduced to the infant immunisation schedule of the Kathmandu valley on 21st August 2015 (see Section 2.1.1). 10-valent PCV was introduced following large earthquakes affecting the Kathmandu Valley in late April 2015. The immediate mortality and morbidity, and displacement of people from their homes associated with the earthquake, was compounded by a severe shortage of clean petroleum fuel (used for vehicles and cooking) from September 2015 to February 2016 (Section 2.1). Within this context, the Inpatient Pneumonia Study sought to investigate the impact of 10-valent PCV on pneumonia in children admitted to Patan Hospital with a clinical diagnosis of pneumonia.

The low prevalence of invasive pneumococcal disease, and the absence of an accurate diagnostic test for pneumococcal pneumonia, limit direct measurements of the impact of PCV on "proven" pneumococcal disease. In Chapter 3 I explored associations between pneumococcal pneumonia with invasive pneumococcal disease (i.e. definite pneumococcal pneumonia) and proxy markers of pneumococcal pneumonia. The proxy markers significantly associated with definite pneumococcal pneumonia in adjusted logistic regression analyses were: radiographic endpoint consolidation (sensitivity 0.94, specificity 0.66, PPV 0.04), plasma CRP concentration $\geq$58 mg/l (sensitivity 0.90, specificity 0.78, PPV 0.08), and NP carriage of pneumococcal serotype 1 (sensitivity 0.17, specificity 0.98, PPV 0.15). CRP concentration was only measured on children enrolled from February 2015 onwards.

Here, I assessed the impact of PCV on the prevalence of pneumonia with endpoint consolidation, in addition to assessing the impact of PCV against clinical phenotypes of childhood pneumonia, as defined by the WHO clinical criteria[15]) and pneumonia without wheeze on clinical examination, in children enrolled to the Inpatient Pneumonia Study. Cases of children with NP carriage of pneumococcal serotype 1 were too few for formal analysis. Instead, I assessed impact of PCV on NP carriage of all vaccine-type pneumococci. Although NP carriage of vaccine-type pneumococci is not a disease *per se*, carriage is thought to be a biological prerequisite for disease[50], and blockage of NP transmission is thought to mediate a large proportion of the effects of PCV on disease[240].

I initially describe data on the numbers of children admitted to Patan Hospital with all-cause pneumonia, clinical phenotypes of pneumonia, endpoint consolidation and NP carriage of vaccine-type pneumococci from March 2014 to December 2016 (outcomes with, respectively, increasing specificity for the measurement of 10-valent PCV impact). I used interrupted time series analyses[23, 241, 242], and restricted the interrupted time series analysis to children <2 years of age since any direct effect of PCV would only be detectable within this age group with this limited (15 month) post-introduction dataset. Given the potential for confounding of time series analyses by the earthquakes of late April 2015, and the fuel shortages of winter 2015–2016, I also used the indirect cohort

method to assess the impact of 10-valent PCV on clinical phenotypes of pneumonia, endpoint consolidation and NP carriage of vaccine-type pneumococci[243, 244]. This method directly assesses odds ratios in "cases" (children with the outcome of interest, e.g. endpoint consolidation) vs "controls" (children without the outcome of interest), and derives a measurement of vaccine effectiveness against each outcome.

4.2 Methods

4.2.1 General methods

Ethics statement

See Chapter 3.

Study site and population

Details of the study site are described in Section 2.1.

Summary of clinical, radiological, laboratory and microbiological methods

The background to studies of childhood pneumonia at Patan Hospital is described in Section 2.3. Definitions, inclusion criteria, exclusion criteria, and clinical methods for the Inpatient Pneumonia Study are described in Section 3.2.3.

In brief, the study commenced enrolment on 20th March 2014, and these data include children enrolled until the end of December 2016. Children $\geq$2 months (60 days) of age and up to 14 years of age who were being admitted to Patan Hospital with a clinical diagnosis of pneumonia were eligible for the study. The clinical diagnosis of pneumonia was made by the admitting paediatric team. Children of consenting parents/guardians were enrolled. All children had a chest radiograph taken at admission/enrolment (analysed according to standardised methods[30]), and a NP swab for culture of pneumococci and Quellung serotyping at Patan Hospital. Blood was sampled in an aseptic manner from all children enrolled to the Inpatient Pneumonia Study for isolation and culture of aerobic bacteria, full blood count and white cell differential at Patan Hospital and retrospective quantitation of plasma CRP concentrations (from February 2015 onwards). Detailed

microbiological and radiological methods are described in Sections 2.2.2, 2.3.2 and 2.3.3.

Data management and analysis

Data were entered and managed on a study-specific database and underwent extensive data checking procedures prior to analysis. Data analysis and figures were produced using *Stata v.14.1* (Statacorp, USA) and R[218], including the package $ggplot2$[245].

Descriptive analyses

Data were initially analysed as tables of counts per (Gregorian) calendar month. Data were analysed across all age groups, before being analysed by the age groups: 2 months to <2 years of age ("<2 years of age" hereafter), 2 years to <5 years of age, and 5 years to 14 years of age. These groups were chosen as follows.

The age group <2 years of age would be the most sensitive to impact of 10-valent PCV, given that children in this age group are overwhelmingly the only children to have received 10-valent PCV. The age group 2 to <5 years of age represents other children to whom the WHO clinical criteria for the diagnosis of pneumonia (Table 1.1) are applicable, and who have typically been enrolled in pneumonia aetiology studies[10, 246]. Enrolment of older children (5 to 14 years of age) provided an important group for the detection of invasive pneumococcal disease in children with a clinical diagnosis of pneumonia (see Chapter 3).

4.2.2 Interrupted time series analyses

The interrupted time series methods used regression models based on that described by Lopez Bernal et al.[241]. A scaling adjustment was made to the Poisson regression model to account for over-dispersion in these low samples sizes, leading to a quasi-Poisson distribution (specifically, a scale parameter estimated by the Pearson χ^2 statistic divided by the residual degrees of freedom[242]). Two or more doses of 10-valent PCV was considered "vaccinated" for this under two years age group[247], with children receiving <2 doses considered "unvaccinated". Introduction of 10-valent PCV – i.e. the "interruption" – was modelled as a continuous proportion of children enrolled to the

Inpatient Pneumonia Study who were aged <2 years and vaccinated with two or more doses of 10-valent PCV (i.e. considered "vaccinated").

I used monthly counts of enrolment of children with "fast breathing", "chest indrawing" and "danger signs", and pneumonia with absence of wheeze and endpoint consolidation to assess the impact of 10-valent PCV on clinical pneumonia in children <2 years of age over time. To account for changes in health-seeking behaviour and seasonal variation in admissions to hospital beyond those attributable to pneumonia I used monthly counts of children admitted to hospital without a diagnosis of pneumonia as a denominator for these outcome variables. Thus the modelled clinical and radiographic outcomes entered to interrupted time series analyses were odds ratios for admission with the specified pneumonia outcome vs admission without a diagnosis of pneumonia (Table 4.1).

For the interrupted time series analysis of NP carriage of vaccine-type pneumococci I used quarterly results (i.e. over three month periods, with the first period of each year beginning 1st January) due to counts of zero in some months. I used counts of NP carriage of non-vaccine-type pneumococci in children with pneumonia as the denominator, generating odds ratios for carriage of vaccine-type vs non-vaccine-type pneumococci in pneumonia. Children with NP carriage of non-vaccine type pneumococci were chosen as denominators (controls) since transmission and NP carriage of non-vaccine type pneumococci were considered to be equally affected by (non-vaccine) secular changes in the community. I hypothesised that such changes might include crowding (anecdotally increased following the earthquakes of 2015), increased ambient pollution (anecdotally increased during the fuel shortage and beyond due to the use of poor quality fuels), and any changes in antibiotic use (both vaccine-type and non-vaccine-type pneumococci have generally similar antimicrobial resistance profiles in Patan, data not shown). This also is consistent with the use of the indirect cohort method to assess impact of pneumococcal conjugate vaccines on NP carriage [244] and invasive pneumococcal disease [101, 243], described below.

The interrupted time series analysis used the following three variables:

Table 4.1: Outcomes variables for interrupted time series analyses

Outcome of interest	Numerator	Denominator
"Fast breathing" pneumonia	"Fast breathing" pneumonia/month	Non-pneumonia admissions/month
"Chest indrawing" pneumonia	"Chest indrawing" pneumonia/month	Non-pneumonia admissions/month
Pneumonia with "danger signs"	Pneumonia with "danger signs"/month	Non-pneumonia admissions/month
Pneumonia without wheeze	Pneumonia without wheeze/month	Non-pneumonia admissions/month
Radiographic endpoint consolidation	Radiographic endpoint consolidation/month	Non-pneumonia admissions/month
Carriage of VT pneumococci	Carriage of VT pneumococci/month	Carriage of non-VT pneumococci/month

i. T: the time elapsed since the start of the study (here, from March 2014 to December 2016),

ii. X_t: a dummy variable indicating the pre-intervention period, and the post-intervention period (in this analysis this was the proportion of children with the outcome variable who had received two or more doses of 10-valent PCV)

iii. Y_t: the outcome variable at time t.

I used these variables to create the segmented regression model:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t \tag{4.1}$$

where β_0 represents the baseline of the outcome at the beginning of the time series (i.e. $T = 0$), β_1 represents the baseline of the outcome in the pre-intervention period, β_2 is the level change in the outcome following the intervention, and β_3 is the slope change following the intervention. β_3 uses the interaction between time and intervention, TX_t. Inspection of the data in the 18 months of pre-intervention period showed no seasonal pattern to the modelled outcome variables. Accordingly, I did not account for seasonality in the regression model.

4.2.3 Indirect cohort method

The indirect cohort method assesses vaccine effectiveness by the use of a case-control design, where those with the outcome of interest are treated as "cases", and those without the outcome of interest are "controls", as described above. The data for this analysis were limited to those from children who were eligible for at least two doses of 10-valent PCV by routine infant immunisation (i.e. not through the "catch-up" campaign). These children were born after the 10th July 2015 and were all greater than 10 weeks of age. As for time series analyses, children were deemed to be "vaccinated" if they had received two or more doses of 10-valent PCV, and children who had received fewer than 2 doses were deemed to be "unvaccinated".

In the specific example of assessing vaccine effectiveness (VE) against NP carriage of vaccine-type pneumococci, children with carriage of vaccine-type pneumococci are cases, and children with carriage of non-vaccine-type pneumococci are controls. The fundamental equation is shown below:

$$VE = 1 - \left(\frac{VT_{vaccinated}/NVT_{vaccinated}}{VT_{unvaccinated}/NVT_{unvaccinated}} \right) \tag{4.2}$$

where VT = NP carriage of vaccine-type pneumococci and NVT = NP carriage of non-vaccine-type pneumococci, in vaccinated vs unvaccinated children. For analysis of clinical pneumonia phenotypes, and pneumonia with endpoint consolidation I used denominators as described above (Table 4.1). I adapted the equation above by the use of a generalised linear model to adjust for age and sex of the child. Data were not adjusted for changes in NP carriage of non-vaccine-type pneumococci in the community in this early phase of 10-valent PCV introduction.

4.3 Results

4.3.1 Descriptive analyses of enrolled patients over time

From March 2014 to end December 2016, 938 children were enrolled in the Inpatient Pneumonia Study. Of 938 children enrolled, 611 (65%) children were under 2 years of age, 215 (23%) children were 2 to 5 years of age and 112 (12%) children were 5 to 14 years of age. The characteristics of all children enrolled in the Inpatient Pneumonia Study are explored by age group, presence or absence of wheeze on examination, radiographic result (endpoint consolidation, infiltrates or normal) and NP carriage of pneumococcal serotypes in Chapter 3: Tables 3.3, 3.4, 3.6, 3.7 and Figure 3.7.

Figure 4.1 describes the numbers of children enrolled with clinician diagnosed pneumonia by age group, and non-pneumonia admissions and total admissions to Patan Hospital during the study period. Enrolment of children with pneumonia of all ages remained relatively constant until September 2015, where increased numbers of children with pneumonia presented to the hospital, particularly children <2 years of age (explored in detail later). As described previously (Section 2.1) major earthquakes occurred in the Kathmandu Valley, and local regions, in late April and May 2015 leading to considerable loss of life, damage to property and the formation of internally-displaced people's camps within Kathmandu. 10-valent PCV was introduced into the infant immunisation schedule in the Kathmandu Valley in late August 2015. In addition, the supply of petroleum and cooking gas to the Kathmandu Valley (and elsewhere in Nepal) was severely interrupted from September 2015 to February 2016.

10-valent PCV introduction

The proportion of children enrolled to the Inpatient Pneumonia Study vaccinated with one or more, two or more doses of 10-valent PCV (i.e. were "vaccinated", as defined here) increased steadily following introduction of PCV10 on 21st August 2015. By December 2015, 82% of enrolled children under 2 years of age had received two or more doses. In contrast, very few children >2 years of age had received 10-valent PCV (Figure 4.2).

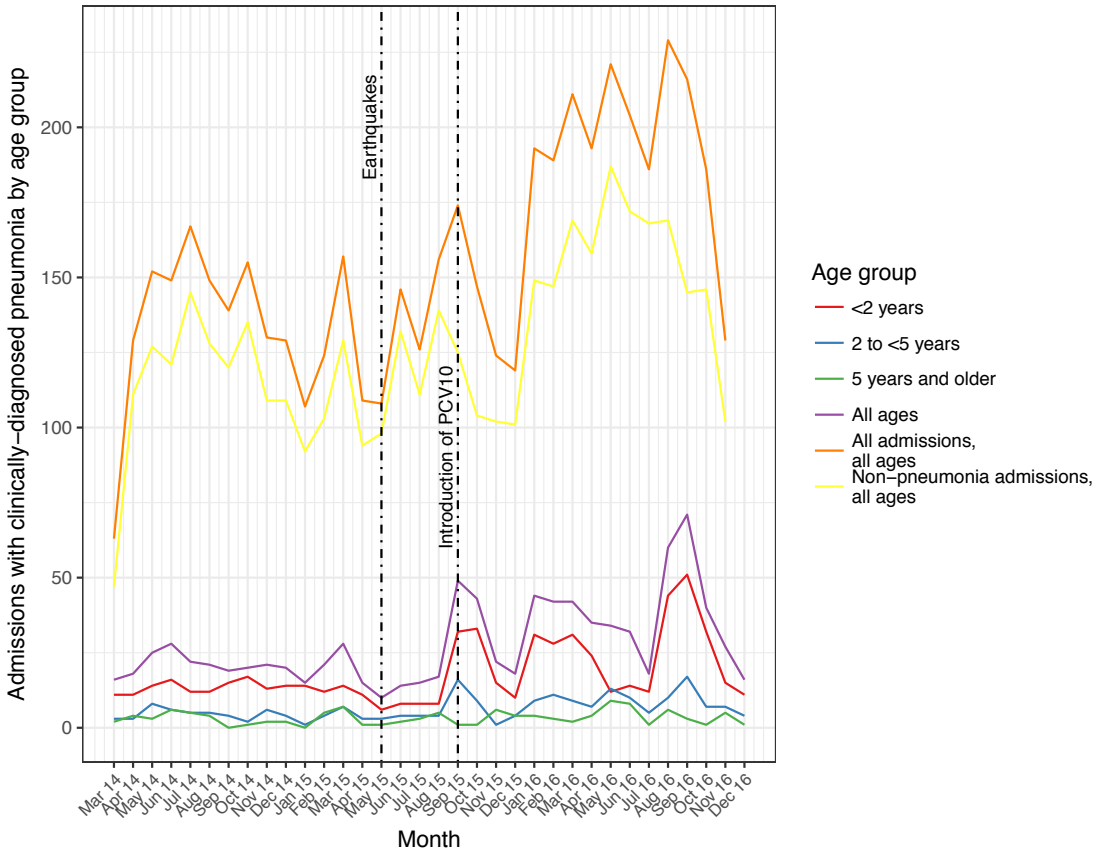


Figure 4.1: Enrolment of children 2 months to 14 years of age with clinician diagnosed pneumonia by age group, by all ages, total non-pneumonia admissions, and all childhood admissions to Patan Hospital per month, March 2014–December 2016. PCV10, 10-valent PCV.

Clinical, laboratory, radiographic and microbiological characteristics

The clinical features of enrolled children by month of enrolment are described in Figure 4.3 and by month and age group (in children <5 years of age) in Figure 4.4. In brief, the apparent increases in enrolment of children with pneumonia (Figure 4.1) is accompanied by apparent increases in all clinical phenotypes of childhood pneumonia, and does not appear to represent an increase in enrolment of a single clinical phenotype only. This interpretation remains following subsetting of the data by age of child (in children <5 years of age; Figure 4.4).

Radiographic endpoint consolidation was predictive of invasive pneumococcal disease (i.e. definite pneumococcal pneumonia) in adjusted logistic regression analyses (see Chapter 3). The radiographic features (as endpoint consolidation and infiltrates) of enrolled children by month of enrolment are described in Figure 4.5, and by month

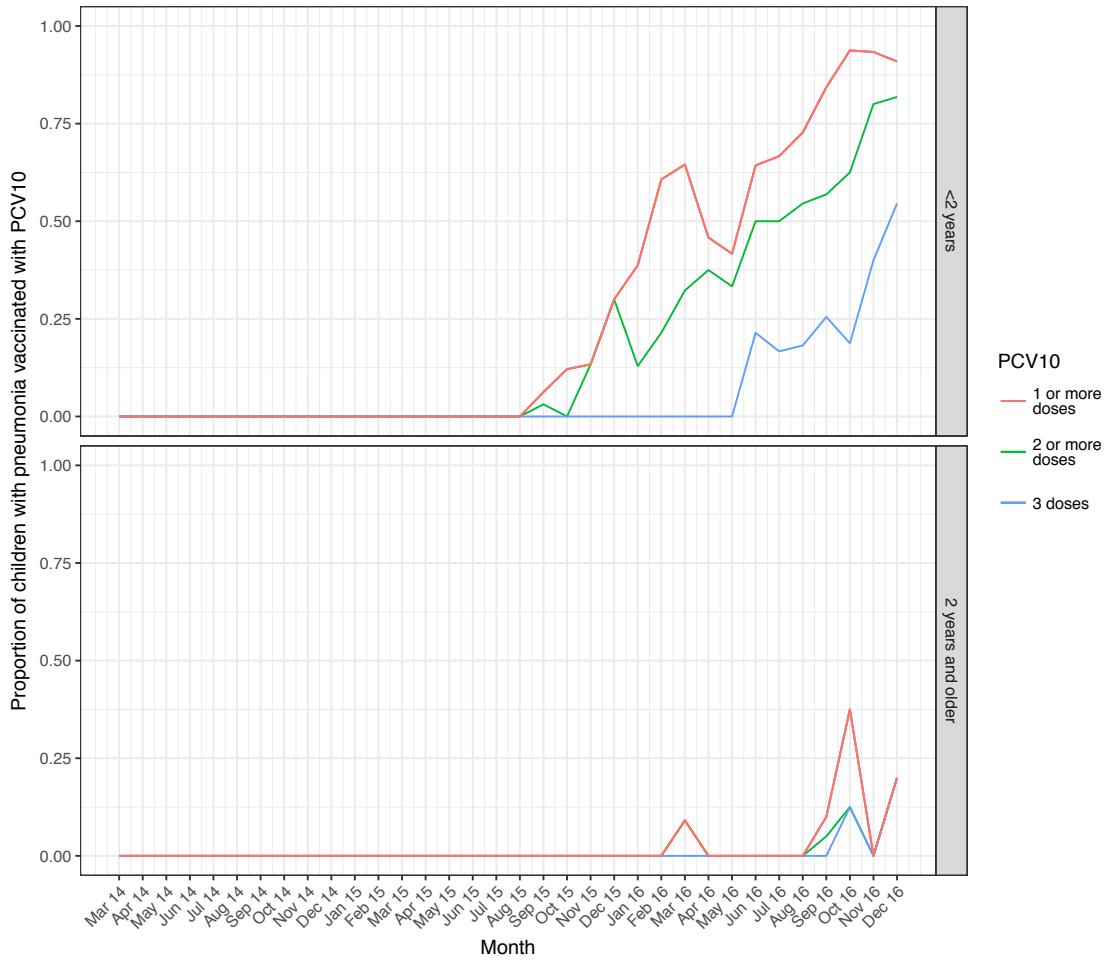


Figure 4.2: The proportion of children vaccinated with one or more, two or more, or three doses of 10-valent PCV, by age group. PCV10, 10-valent PCV.

and age group in Figure **??**. In these descriptive analyses, the sharp increase (and rapid decrease) in pneumonia admissions of September–October 2015 is associated with a high proportion of pneumonia with infiltrates relative to endpoint consolidation in comparison to the preceding months. In contrast, there is a prolonged period from December 2015 to July 2016 where there is a high proportion of pneumonia with endpoint consolidation relative to infiltrates in comparison to the preceding months. These apparent changes in the proportions of pneumonia with infiltrates and endpoint consolidation are most pronounced in children <2 years of age (Figure **??**).

NP carriage of pneumococcal serotypes

The prevalence of NP carriage of pneumococcal serotypes and invasive pneumococcal disease in the case series may be affected by antibiotics prior to sampling. The proportion

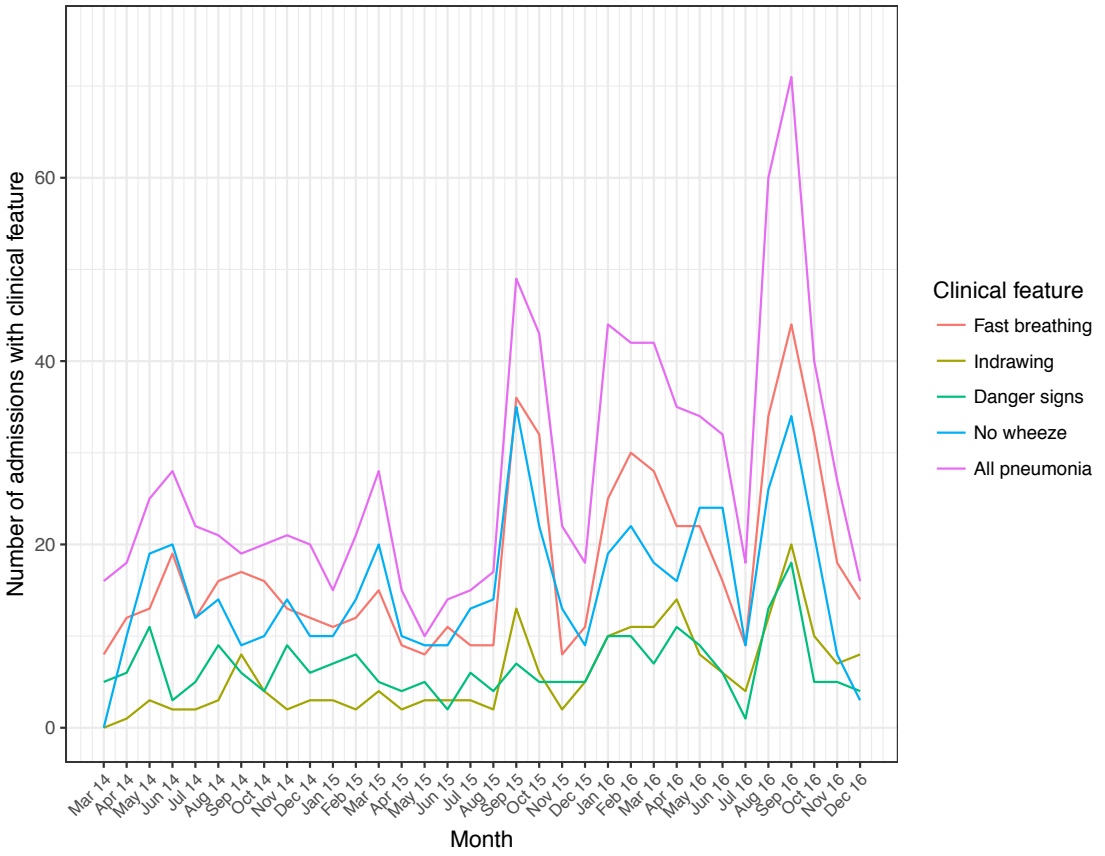


Figure 4.3: Enrolment of children 2 months to14 years of age with clinician-diagnosed pneumonia, according to clinical feature, by month of enrolment, March 2014–December 2016.

of children enrolled in the study with prior antibiotic therapy, relative to those enrolled without prior antibiotic therapy, remained constant over time (χ^2 test of trend, p = 0.758, Figure 4.7). The number of children enrolled in the study with NP carriage of pneumococci grouped by vaccine-type is described by month in Figure 4.8, and by month and age group in Figure 4.9. As with other outcome variables, these data are explored in detail below.

4.3.2 Results of interrupted time series analyses

Table 4.2 shows the results of the quasi-Poisson regression interrupted time series analyses in children <2 years of age in the Inpatient Pneumonia Study enrolled between March 2014 and end December 2016. The "trend" is based upon the change in the outcome over the time period omitting the interruption (the "counter-factual"). The impact

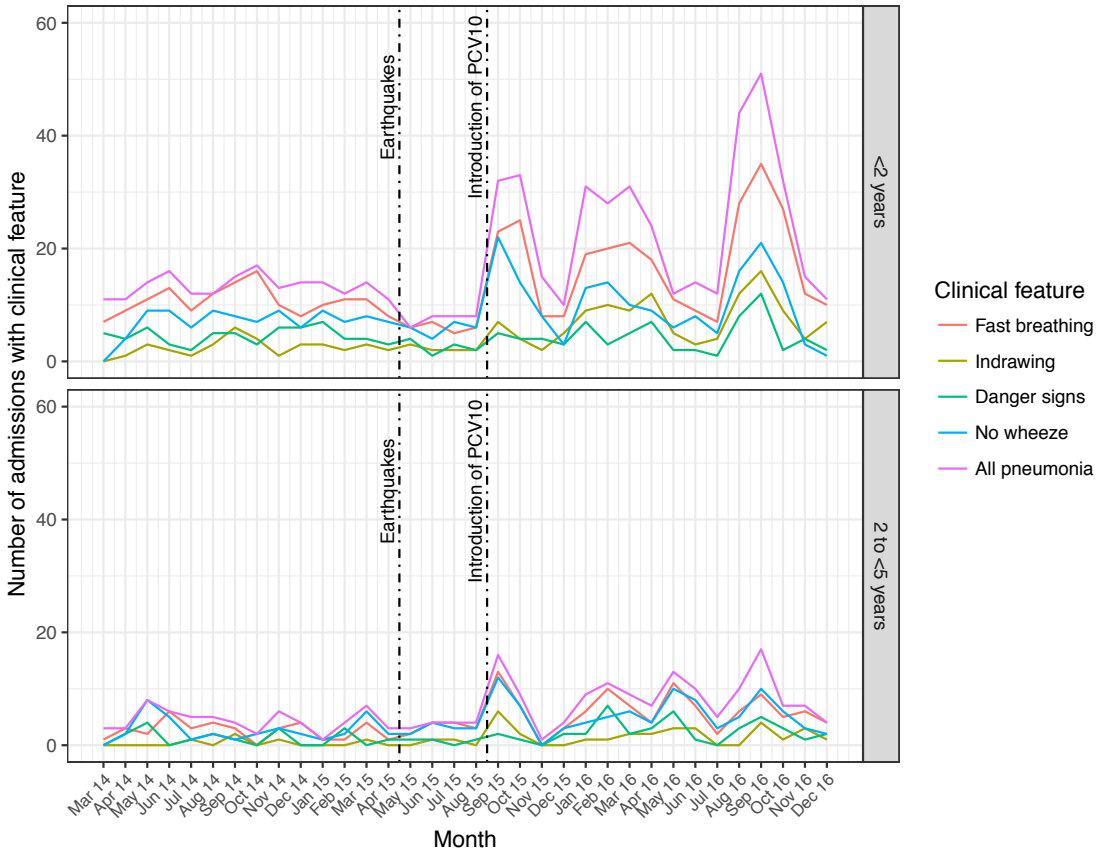


Figure 4.4: Clinical features of enrolled children under 5 years of age admitted to Patan Hospital with clinician diagnosed pneumonia by age group, March 2014–December 2016.

of the interruption, 10-valent PCV introduction as a proportion of children within the case series per month/quarter, is modelled separately. With an increasing proportion of children vaccinated with 10-valent PCV there was no significant change in the proportion of children admitted to Patan Hospital with all pneumonia, or pneumonia with absence of wheeze relative to all cause child admissions to Patan Hospital following August 2015 (Figures 4.10 and 4.11). There was a small significant increase in fast breathing pneumonia (4%, p = 0.002), chest indrawing pneumonia (4%, p = 0.021) and pneumonia with danger signs (4%, p = 0.003) relative to non-pneumonia admissions temporally associated with increasing proportions of children vaccinated with 10-valent PCV. Within this period, endpoint consolidation relative to non-pneumonia admissions (Figure 4.12) and NP carriage of vaccine-type pneumococci relative to non-vaccine pneumococci (Figure 4.13) were not temporally associated with increasing proportions of children vaccinated with 10-valent PCV.

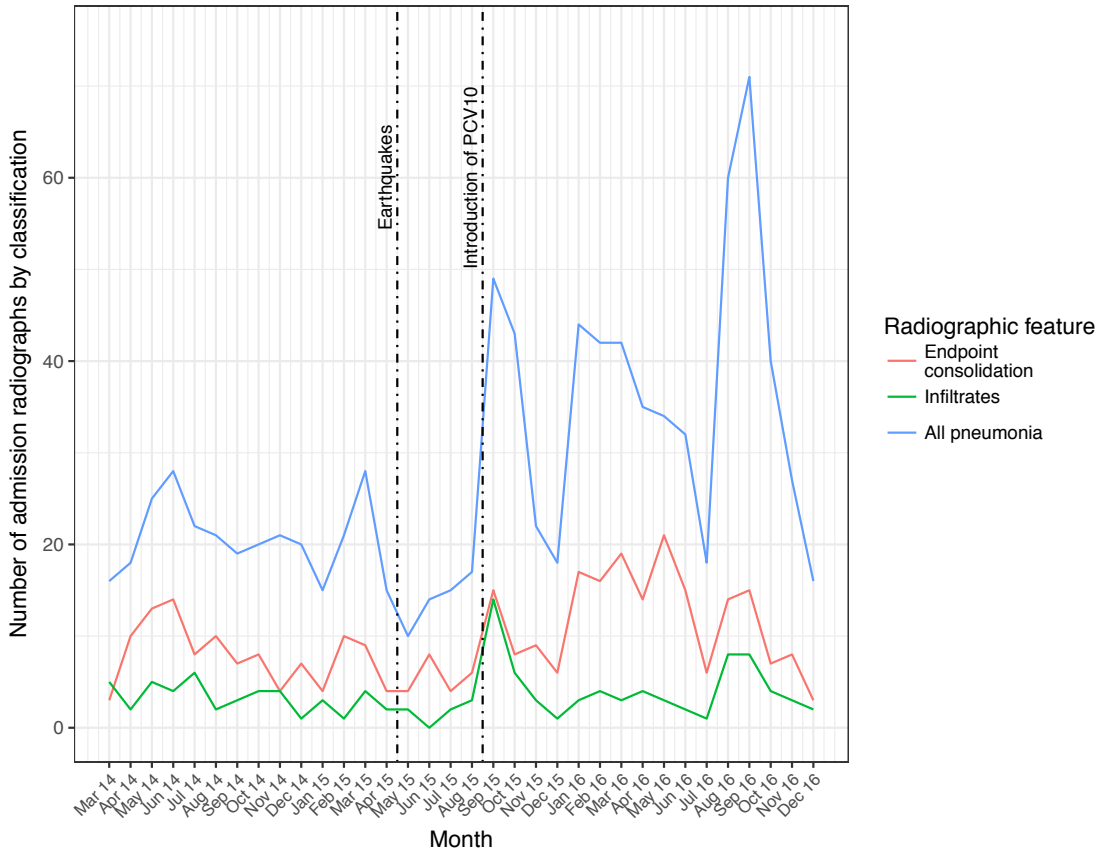


Figure 4.5: Characteristics of children 0 to 14 years of age with clinician-diagnosed pneumonia, according to radiographic classification, by month of enrolment, March 2014–December 2016.

Table 4.2: Interrupted time series analyses of changes in the number of enrolled children under 2 years of age with clinical features (fast breathing, chest indrawing, danger signs, absence of wheeze or radiographic endpoint consolidation / non-pneumonia admissions in this age group), and NP carriage of vaccine-type pneumococci / carriage of non-vaccine-type pneumococci associated with increasing proportion children vaccinated with 10-valent PCV.

	Trend		Impact of 10-valent PCV	
Outcome	**Estimate (95% CI)**	**p**	**Estimate (95% CI)**	**p**
All pneumonia	0.02 (-0.01, 0.06)	0.247	-0.07 (-1.49, 1.34)	0.92
Clinical phenotypes				
Fast breathing	-0.01 (-0.03, 0.01)	0.244	0.04 (0.02, 0.06)	0.002
Chest indrawing	0.02 (-0.01, 0.05)	0.136	0.03 (0.01, 0.05)	0.021
Danger signs	-0.04 (-0.07, -0.02)	0.002	0.04 (0.02, 0.07)	0.003
No wheeze	-0.01 (-0.03, 0.02)	0.582	0.02 (-0.01, 0.05)	0.156
Radiographs				
Endpoint consolidation	0.00 (-0.03, 0.03)	0.895	0.00 (-0.03, 0.03)	0.906
NP carriage				
VT/non-VT pneumococci	-0.02 (-0.29, 0.25)	0.891	-0.34 (-2.13, 1.45)	0.721

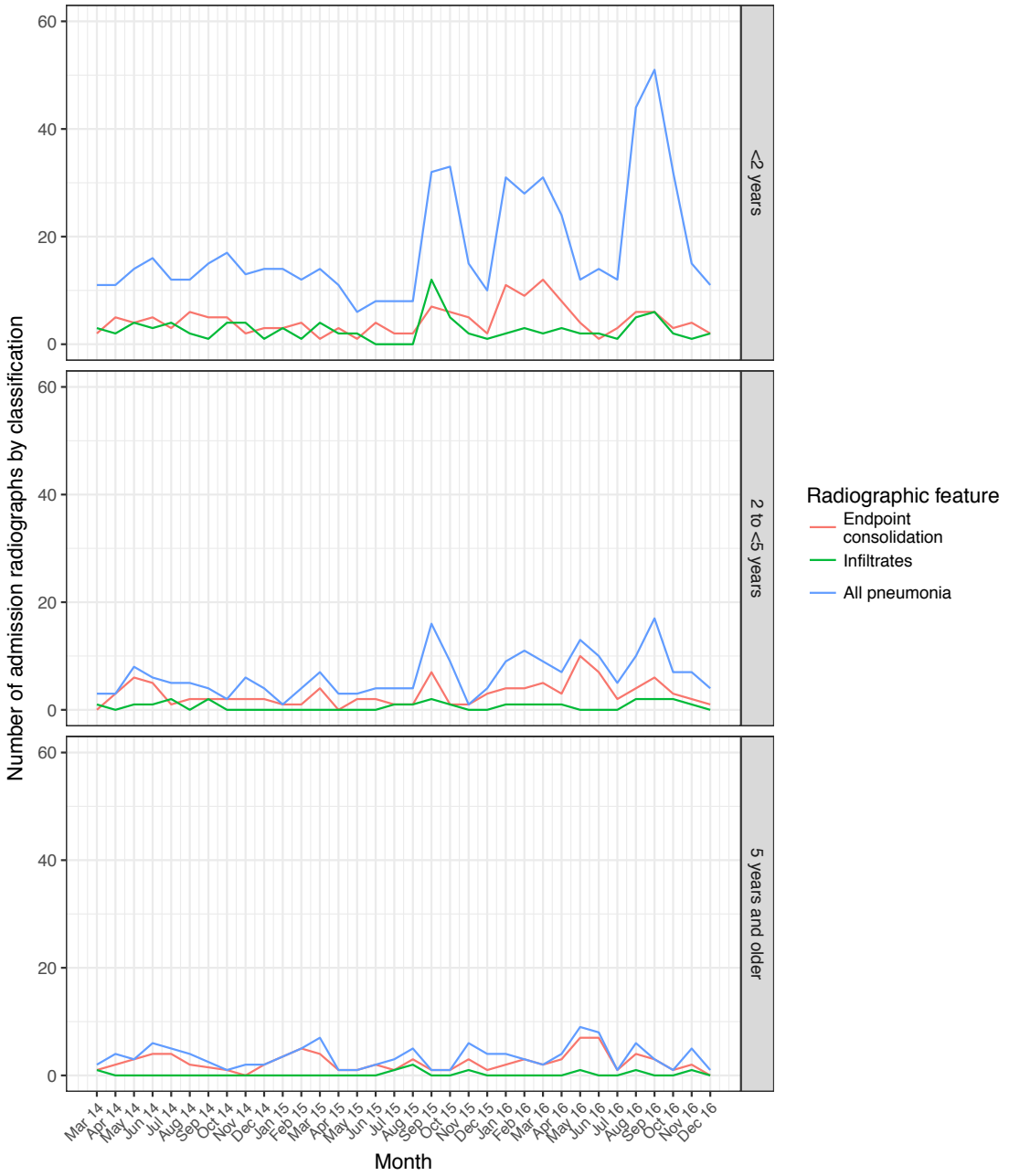


Figure 4.6: Enrolment of children 0 to 14 years of age with clinician-diagnosed pneumonia, according to radiographic feature, by month of enrolment and by age group, March 2014–December 2016.

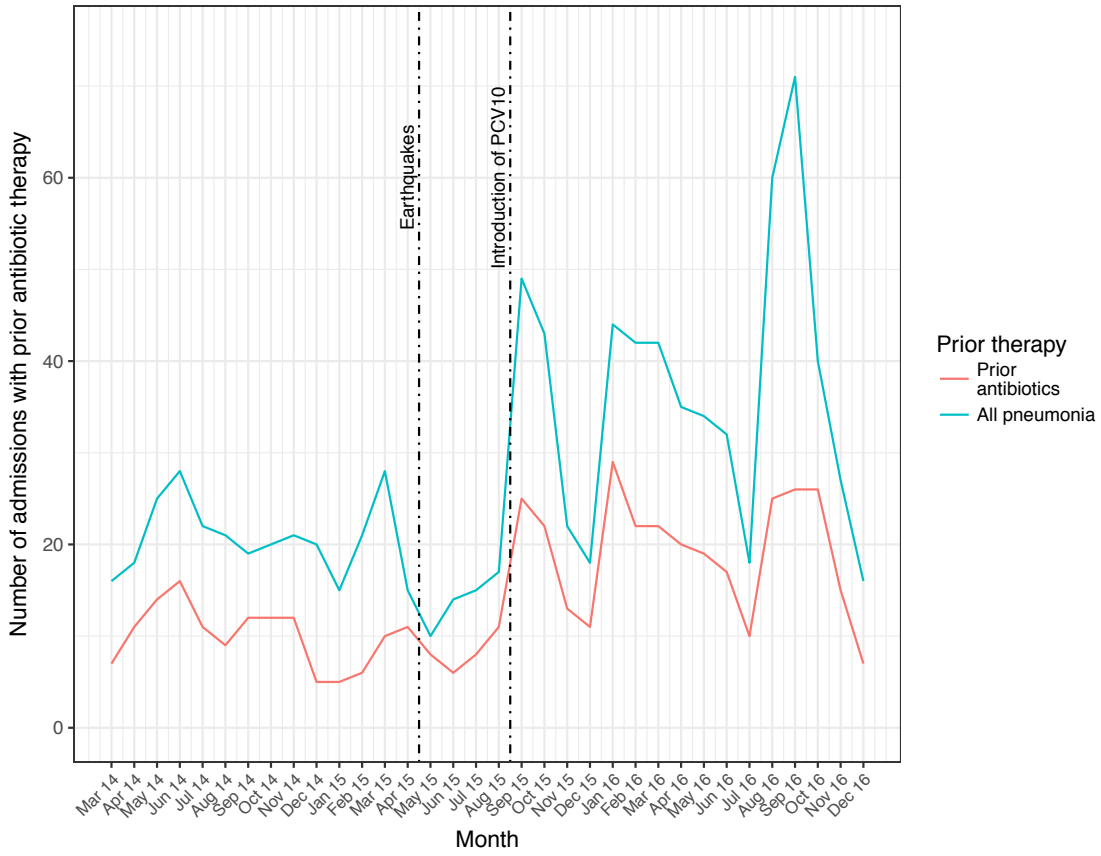


Figure 4.7: Enrolment of children 0 to 14 years of age with clinician-diagnosed pneumonia, according to prior treatment with antibiotics, by month of enrolment, March 2014–December 2016.

4.3.3 Results of indirect cohort methods

The birth cohort of children eligible for at least two doses of 10-valent PCV consisted of 178 children. These children were all aged $\geq$10 weeks of age and born after the 10th July 2015 (i.e. were unborn or <6 weeks of age on 21st August 2015 when PCV was introduced to the Kathmandu Valley). Children who received two doses of 10-valent PCV were significantly older than children who had not received two doses of the vaccine (median age 6.9 months, IQR 5.3, 9.8, vs 3.5 months, IQR 2.6, 6.7; $p < 0.001$). Other characteristics of these patients are shown in Table 4.3.

The results for the analysis of vaccine effectiveness using the indirect cohort method are shown in Table 4.4. In adjusted regression analyses, there was no significant effect of 10-valent PCV on fast breathing pneumonia, chest indrawing pneumonia, pneumonia with danger signs, or pneumonia with endpoint consolidation. In contrast, vaccine

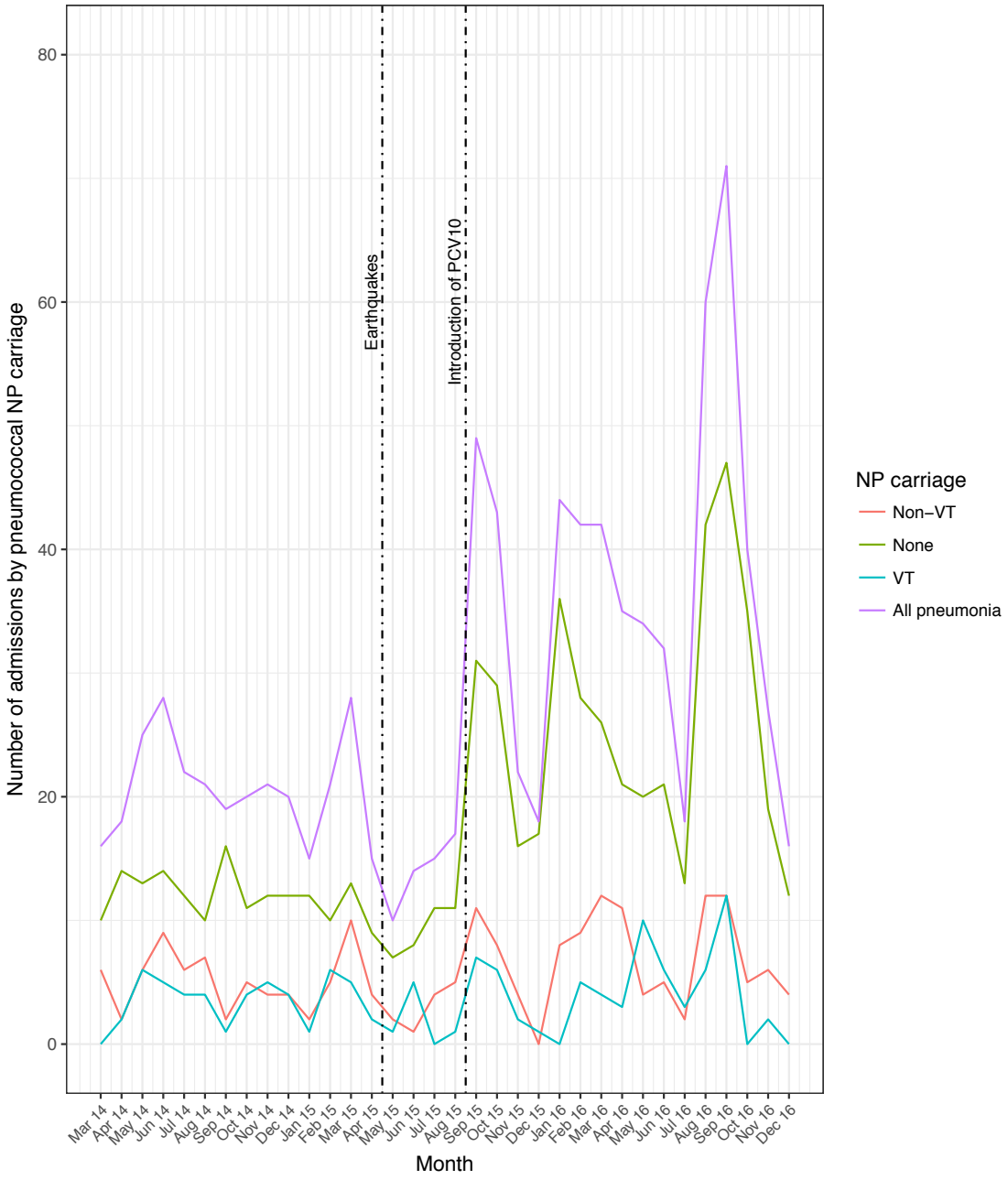


Figure 4.8: Enrolment of children 0 to 14 years of age with clinician diagnosed pneumonia, according to carriage of pneumococcal serotypes, by month of enrolment and age, March 2014–December 2016. Non-VT, non-vaccine-type pneumococci; VT, vaccine-type pneumococci.

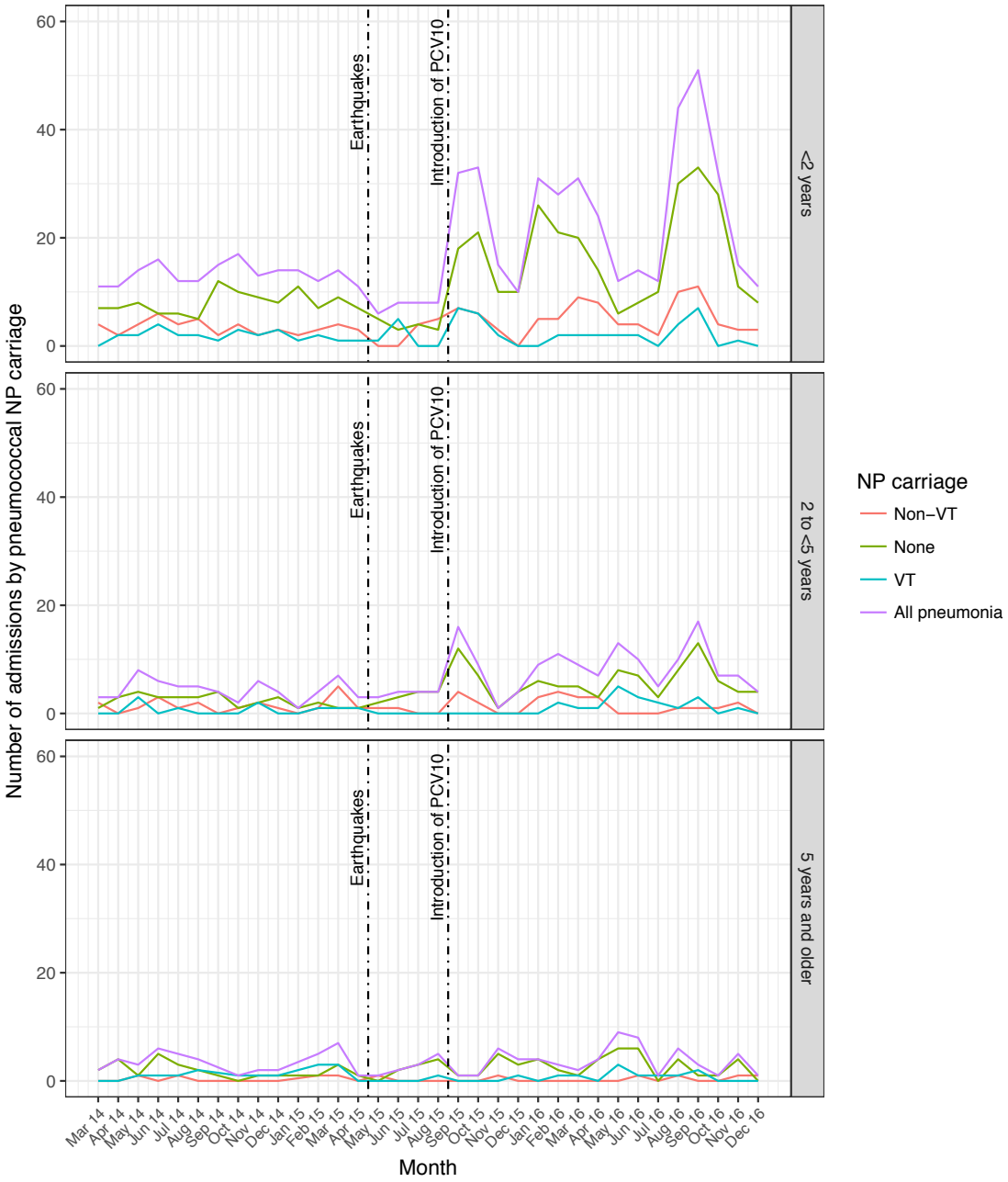


Figure 4.9: Enrolment of children 0 to 14 years of age with clinician diagnosed pneumonia, according to NP carriage of pneumococcal serotypes, by month of enrolment and age, March 2014–December 2016. Non-VT, non-vaccine-type pneumococci; VT, vaccine-type pneumococci.

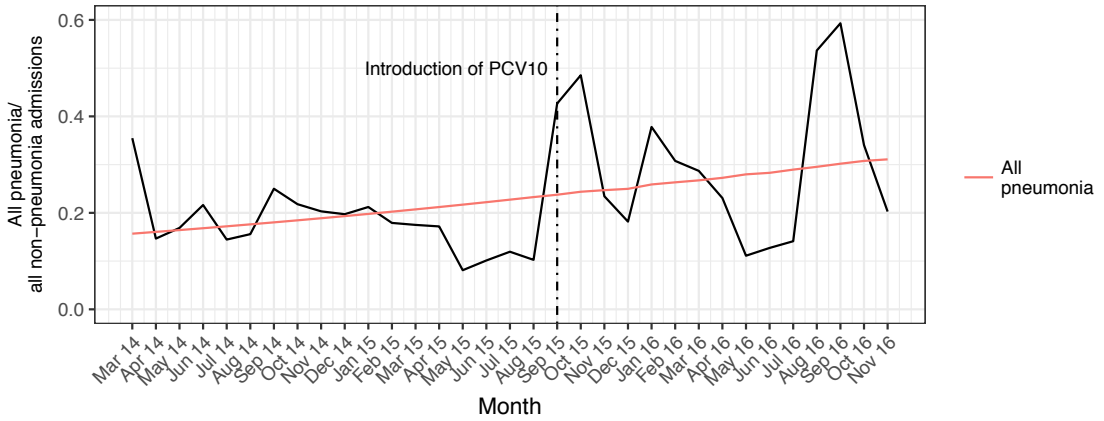


Figure 4.10: Analysis of all pneumonia admissions by month of enrolment in children <2 years of age in the Inpatient Pneumonia Study. Black line: measured outcome in children <2 years of age / monthly non-pneumonia admissions in children <2 years of age. Red line: modelled outcome, adjusted for increasing coverage of two or more doses of 10-valent PCV in this age group.

effectiveness (as percentage) against NP carriage of vaccine-type pneumococci was 84% (95% CI 1%, 98%) although this result was of marginal significance (p = 0.055).

4.4 Discussion

4.4.1 Interrupted time series analyses

Interrupted time series analyses showed a significant increase of 4% in the proportion of children with pneumonia and fast breathing, a 3% increase in children with pneumonia and chest indrawing, and a 4% increase of pneumonia with danger signs temporally associated with introduction of 10-valent PCV. In this analysis, there were no significant effects of 10-valent PCV against pneumonia with endpoint consolidation and NP carriage of vaccine-type pneumococci. These data are in contrast to a number of other analyses showing decreases in childhood clinical pneumonia and pneumonia with endpoint consolidation following introduction of PCV in the USA[248, 249], South America[250, 251], Australia[252] and Israel[253]. This may be related to the short post-introduction period of data analysed here – or the widespread changes in pneumonia epidemiology expected following the major earthquakes of April and May 2015 (discussed below).

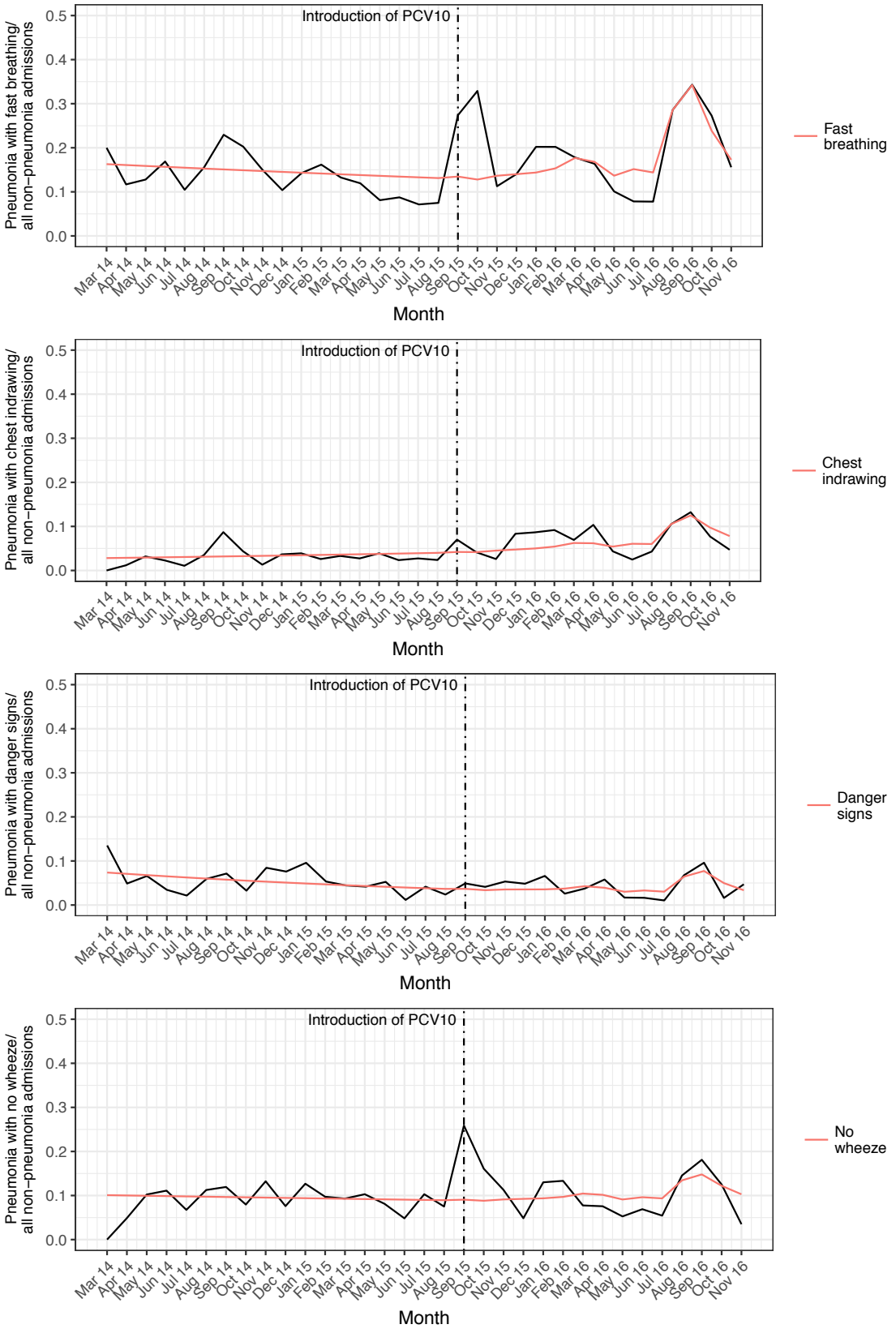


Figure 4.11: Analysis of clinical pneumonia phenotypes by month of enrolment in children <2 years of age in the Inpatient Pneumonia Study. Black line: measured outcome in children <2 years of age / monthly non-pneumonia admissions in children <2 years of age. Red line: modelled outcome, adjusted for increasing coverage of two or more doses of 10-valent PCV in this age group.

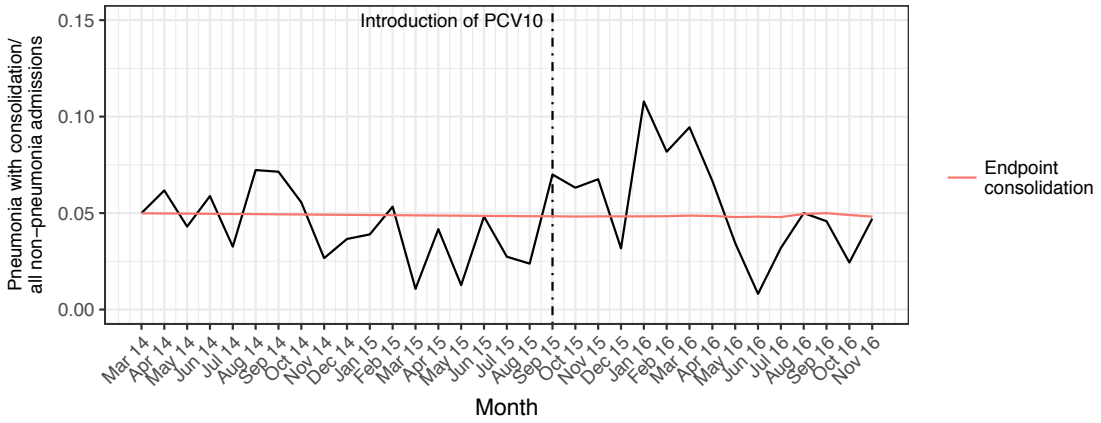


Figure 4.12: Analysis of pneumonia with endpoint consolidation by month of enrolment in children <2 years of age in the Inpatient Pneumonia Study. Black line: measured outcome in children <2 years of age / monthly non-pneumonia admissions in children <2 years of age. Red line: modelled outcome, adjusted for increasing coverage of two or more doses of 10-valent PCV in this age group.

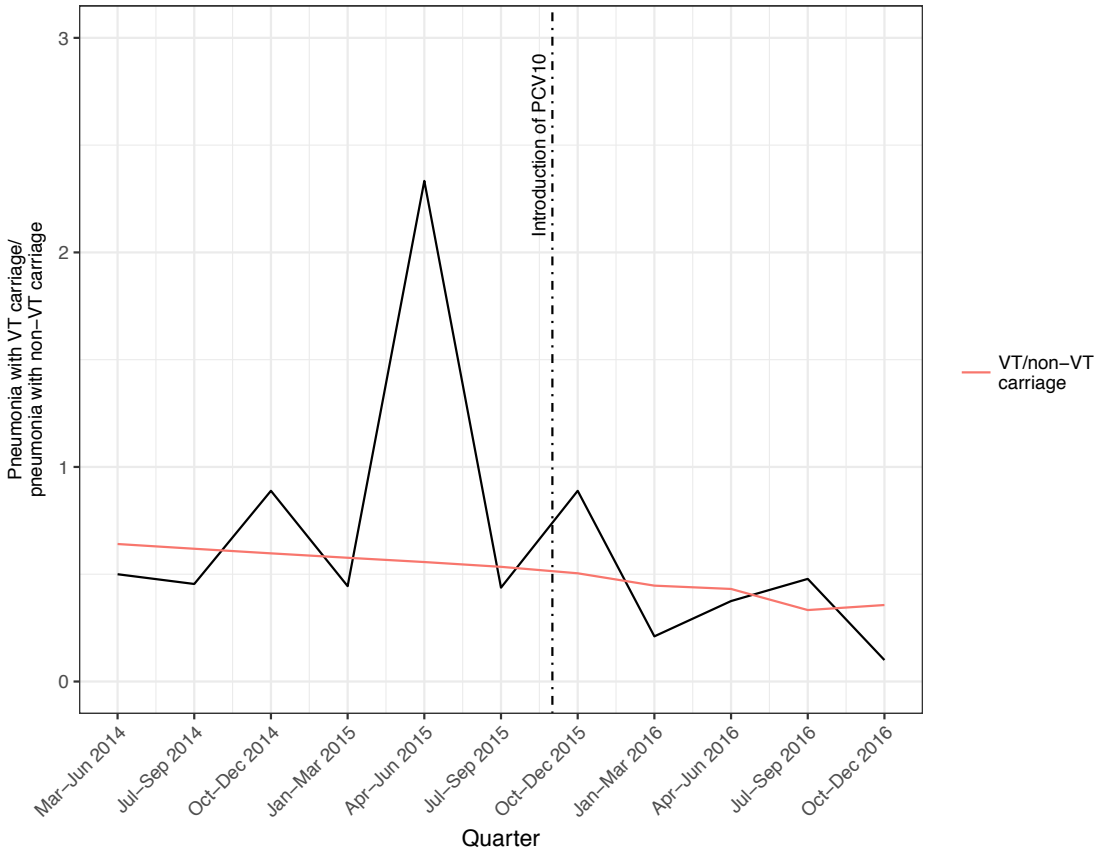


Figure 4.13: Analysis of NP carriage of vaccine-type pneumococci / non-vaccine-type pneumococci over time in children <2 years of age in the Inpatient Pneumonia Study. Black line: measured outcome in children <2 years of age / quarterly non-pneumonia admissions in children <2 years of age. Red line: modelled outcome, adjusted for increasing coverage of two or more doses of 10-valent PCV into this age group. VT, vaccine-type.

Table 4.3: Characteristics of patients enrolled in the study, and eligible for two or more doses of 10-valent PCV through routine immunisation, by vaccination status. The analyses of vaccine effectiveness using the indirect cohort method was done using data from this subset of the case series.

	10-valent PCV vaccination status			
	Two or more doses ("vaccinated")	One or fewer doses ("unvaccinated")	Total	p[a]
Total	124 (70%)	54 (30%)	178	
Male sex	85 (69%)	34 (63%)	119 (67%)	0.579
Age				
Age in months (median, IQR)	6.9 (5.3, 9.8)	3.5 (2.6, 6.7)	6.2 (4.3, 9.2)	<0.001
Number of doses of 10-valent PCV				
No doses	0	23 (43%)		
One dose	0	31 (57%)		
Two doses	89 (72%)	0		
Three doses	35 (28%)	0		
Clinical features				
Fast breathing pneumonia	92 (74%)	40 (74%)	132 (74%)	1
Chest indrawing pneumonia	41 (33%)	26 (48%)	67 (38%)	0.082
Chest indrawing/fast breathing pneumonia	95 (77%)	43 (80%)	138 (78%)	0.804
General danger signs[b]	19 (15%)	9 (17%)	28 (16%)	0.998
Wheeze on examination	83 (67%)	33 (61%)	116 (65%)	0.563
Laboratory results				
CRP (median, IQR)	9.2 (2.4, 30)	7.4 (2.1, 15)	8.2 (2.3, 24)	0.684
Neutrophil count x 10^3/mm^3 (median, IQR)	6.0 (4.1, 9.5)	4.5 (3.2, 6.5)	5.5 (3.7, 8.9)	0.076
Radiographic result				
Endpoint consolidation	29 (23%)	10 (19%)	39 (22%)	0.600
Infiltrates	10 (8.1%)	4 (7.4%)	14 (7.9%)	1
No EPC, uninterpretable for infiltrates	0	2 (3.7%)	2 (1.1%)	
Normal	83 (67%)	38 (70%)	121 (68%)	
Uninterpretable, not done	2 (1.6%)	0	2 (1.1%)	
NP carriage of pneumococcal serotypes				
Prior antibiotic therapy	65 (52%)	28 (52%)	93 (52%)	1
Carriage of any pneumococci	35 (28%)	13 (24%)	48 (27%)	0.700
Carriage of any pneumococci (no prior antibiotics)	22/59 (37%)	10/26 (39%)	32/85 (37%)	1
PCV10 pneumococci	5 (4.0%)	4 (7.4%)	9 (5.1%)	0.567
Additional PCV13 pneumococci	2 (1.6%)	1 (1.9%)	3 (1.7%)	1
Non-PCV pneumococci	22 (18%)	6 (11%)	28 (16%)	0.372
Non-typeable pneumococci	6 (4.8%)	2 (3.7%)	8 (4.5%)	1
Invasive bacterial disease				
Invasive pneumococcal disease	0	0	0	1
Other invasive pathogens[c]	2 (1.6%)	2 (3.7%)	4 (2.2%)	0.753

[a] p values calculated using χ^2 test without continuity corrections for two groups as necessary, or using a Welch two-sample t-test of $\log_{10}$ transformed continuous variables (age, CRP and neutrophils count).

[b] Danger signs: unable to drink, persistent vomiting, convulsions, lethargy or reduced consciousness, stridor in a calm child or severe malnutrition (weight-for-age z-score <-3).

[c] In vaccinated and unvaccinated, respectively: *Acinetobacter* spp. 0, 1; *Escherichia coli* 1, 0; *Neisseria meningitidis* 1, 0; *Pseudomonas* spp. 0, 1

Table 4.4: Indirect cohort method: estimates of vaccine effectiveness against various outcome measures in children eligible for two or more doses of 10-valent PCV.

Endpoint	$n_{vaccinated}$	$n_{unvaccinated}$	VE[a] (unadjusted, 95% CI)	p (unadjusted)	VE (adjusted[b], 95% CI)	p (adjusted)
All pneumonia	124	54	-	-	-	-
Clinical phenotypes						
Fast breathing	92	40	-0.01 (-1.06, 0.53)	0.987	-0.11 (-1.46, 0.51)	0.805
Chest indrawing	41	26	0.47 (-0.02, 0.72)	0.058	0.37 (-0.29, 0.70)	0.204
Danger signs	19	9	0.10 (-1.24, 0.61)	0.821	0.01 (-1.66, 0.61)	0.982
Radiographs						
Endpoint consolidation	29	10	-0.34 (-2.11, 0.38)	0.471	-0.45 (-2.62, 0.38)	0.402
NP carriage						
Any pneumococci/No carriage	35	13	-0.24 (-1.66, 0.40)	0.566	-0.06 (-1.44, 0.53)	0.887
VT/non-VT pneumococci	5	4	0.63 (-0.79, 0.92)	0.203	0.84 (0.01, 0.98)	0.055

[a]VE, vaccine effectiveness

[b]Adjusted for age and sex of the child in logistic regression models.

A recent large-scale time series analysis of childhood pneumonia in Malawi following introduction of 13-valent PCV, showed a significant increase in fast breathing pneumonia, although decreases in pneumonia with hypoxaemia (not systematically measured in our study) and pneumonia-related mortality in children <5 years of age. Given the lack of specificity of fast breathing pneumonia for pneumococcal pneumonia, one could hypothesise that an increase in non-pneumococcal pneumonia might be the underlying cause of the increase in fast breathing pneumonia seen following vaccine introduction in Malawi. Adjustment of modelled data from the Inpatient Pneumonia Study models for respiratory virus infection (particularly respiratory syncytial virus) may be of importance[253].

The time series analyses used a quasi-Poisson regression model, appropriate for count data with small numbers in some time periods. In the absence of population data from which an incidence rate could be calculated, I used a denominator (counts of children admitted without pneumonia) to help account for changes in healthcare utilisation in the Patan community when modelling clinical pneumonia and pneumonia with endpoint consolidation. An alternative approach would be to use a "control" condition[249], or to model a synthetic control condition[254], based on admissions of children with a disease that would not be affected the intervention (10-valent PCV). I took this approach when assessing changes in NP carriage of vaccine-type pneumococci, by the use of non-vaccine-type pneumococci, on the assumption antigens in 10-valent PCV would not have cross-reactivity to non-vaccine antigens. Analysis of data from the case series shows that this assumption is valid, excepting a non-significant decrease in NP carriage of pneumococcal serotype 6A in unadjusted analyses (data not shown).

The short length of data both prior to, and post, introduction of 10-valent PCV limited these analyses – leading to wide confidence intervals on estimates of the changes in most outcomes. In addition, longer term analyses of data from Patan Hospital (Brian Wahl, *personal communication*) show clear seasonal increases in the numbers of children diagnosed with pneumonia between September and December (autumn) annually, excepting in 2014 (the only autumn included in the dataset analysed here). Thus, I was unable to adjust for seasonality. Given that there were two autumns post-introduction

of 10-valent PCV this is likely to be an important residual confounder. The use of the natural log of the outcome variable (i.e. $l_n(Y)_t$) limits extreme values and may represent an addition to the time series analysis used here.

The modelling of the "interruption" as a proportion of children who have received two doses of 10-valent PCV is an oversimplification. A full course of 10-valent PCV consists of 3 doses (in the Nepali schedule), and this may be a more appropriate measure of "vaccination status". These time series models do not take account of individual variation in vaccination – for example, one could hypothesise that increases in phenotypes of clinical pneumonia may be only in unvaccinated children. I address this issue using the indirect cohort method (below).

The outstanding limitations to these analyses are the large and pervasive changes that are likely to have resulted from the earthquakes of April 2015[209]. From personal observation, the earthquakes led to large variations in demographics (as families moved to worse-damaged areas in the surrounding hills and mountains), the formation of internally-displaced peoples camps in Kathmandu (with consequential crowding) and disruption to medical services in the short and medium term. Although data on the impact of crises on the incidence of pneumonia in children are sparse[212], one might hypothesise an increase due to increased transmission of viral and bacterial pathogens in crowded populations. In addition, food and shelter insecurity may increase susceptibility of children to pathogens once acquired.

In addition, political unrest at the border of India and Nepal limited transport of all goods into Nepal from India across this crucial land border from September 2015 to February 2016[210]. Official data are not available for increases in fuel prices, but anecdotally, the black market price of a cylinder of liquid petroleum gas increased tenfold, from US$ 10 to US$ 100. This led to a large increase in burning of low quality fuel during these winter months, with notable increases in ambient air pollution[211], and probably indoor air pollution. Case-control epidemiological studies have consistently shown a relationship between indoor air pollution and large increases in odds of a child in the household developing pneumonia[255], although randomised controlled trials of clean

cooking stoves have given equivocal results of benefit[256, 257]. Time series analyses have shown significant increases in attendance to emergency departments associated with relatively small increases (from one quartile to another) of ambient concentrations of carbon particulates in the city of Atlanta (GA, USA) for children <5 years of age[258]. A recent meta-analysis data from the Global Burden of Diseases Study 2015 estimated that carbon particulates was the fifth-ranking mortality risk-factor globally with 675 000 deaths accounted for by this risk factor annually (of all ages)[16].

Understandably, data for the population movements and crowding that occurred in the weeks and months following the earthquakes of April 2015 do not exist – it is therefore not possible to accurately adjust time series models for these factors. There are no detailed published data on concentrations of ambient air pollutants from the Kathmandu Valley that could be used to adjust time series models appropriately[259], and no proxy data, such as fuel prices, since these were controlled by the Government of Nepal rather than subject to supply and demand. In addition, the close temporal association of these events with introduction of 10-valent PCV, and their likely pervasive and medium–long term effects on child respiratory health in the Kathmandu Valley, precluded modelling separate "interruptions". Thus, the lack of data to estimate the impact of earthquakes on childhood pneumonia, and the lack of data on ambient pollution in the Kathmandu Valley makes the use of any term to adjust for these confounders highly inaccurate; whilst the near contemporaneous timing of earthquakes, vaccine introduction (and increasing coverage) and fuel shortages makes post hoc estimations of the impact of these variables on childhood pneumonia impossible. There is therefore great need for further post-introduction data to provide a full assessment of the impact of 10-valent PCV introduction on childhood pneumonia at Patan Hospital beyond the immediate crises of the earthquakes of 2015, and the fuel shortages of winter 2015–2016. To that end, enrolment to the Inpatient Pneumonia Study is currently ongoing.

4.4.2 Indirect cohort method

Given the likely extent of confounding in the interrupted time series analyses, it was appropriate to use an alternative method to assess the impact of the introduction of 10-valent PCV to the infant immunisation schedule. The indirect cohort method uses the odds of vaccination in a "case" (e.g. a child with pneumonia and NP carriage of vaccine-type pneumococci) vs the odds of vaccination in a "control" (e.g. a child with pneumonia and NP carriage of non-vaccine-type pneumococci; Equation 4.2) to derive an estimate of vaccine effectiveness. Among the 178 children enrolled to the Inpatient Pneumonia Study who were eligible for two or more doses of 10-valent PCV, there was no significant vaccine effectiveness against phenotypes of clinical pneumonia, and pneumonia with endpoint consolidation. Although there was no effectiveness against NP carriage of pneumococci (as a whole), there was a large effect (84%) against the considerably more specific endpoint of NP carriage of vaccine-type pneumococci.

There are specific strengths to this method in this context. Firstly, the method uses individual vaccination data. As noted above, the use of the *proportion* of children vaccinated in the interrupted time series analysis may mask important variation between children who are, and children who are not, vaccinated. Thus, this potential source of confounding is removed. In addition, the indirect cohort method is not reliant on a period of observation prior to the introduction of 10-valent PCV; an important factor given the limited data used here.

A major limitation to this method is splitting of children into "unvaccinated" and "vaccinated" by the whether two or more doses of 10-valent PCV have been received. This binary definition does not account for children who have had one dose of 10-valent PCV (who may have partial protection against NP pneumococcal carriage and pneumococcal pneumonia); or that only with three doses of 10-valent PCV are sustained antibody concentrations above a protective threshold achieved[260]. The small numbers of children in this analysis precluded analysis at this level of detail. If one dose of 10-valent PCV provides some protection, the binary analysis of vaccinated vs unvaccinated presented

here is therefore likely to bias the results towards a null effect of the vaccine. Only 23/178 (13%) children in the cohort had not received a single dose of 10-valent PCV.

Children who received two or more doses of 10-valent PCV ("vaccinated") may differ systematically from children who were "unvaccinated". Children who were vaccinated were significantly older (median age 6.9 months) than children who were unvaccinated (median age 3.5 months) – which may have effects on pneumonia epidemiology independent of vaccination status. I therefore incorporated age into the modelled results. Data not presented here show an increase in NP carriage of vaccine-type pneumococci with age prior to the introduction of 10-valent PCV – this this may also bias the results towards a null effect of the vaccine. In contrast, in the context of Nepal, children who received two timely doses of may be more likely to be from urban families with a higher level of parental education (and socio-economic status). This "healthy-user" effect may increase the attributed vaccine effectiveness – although there is a high level of vaccine uptake generally in Nepal, which is seen across socioeconomic groups (see Section 2.1).

A final limitation to the indirect cohort method is secondary to the hypothesised indirect ("herd") effects of 10-valent PCV on pneumonia, and carriage of vaccine-type pneumococci in children. Children who have not been vaccinated may receive benefits of vaccination from their vaccinated peers due to reduced transmission of vaccine-type pneumococci in the community; thus biasing the results towards a null effect of the vaccine. A method to estimate this bias has been proposed[244], although I did not apply this method here, given probable limited indirect effects of 10-valent PCV in the short period of data post-vaccine introduction.

4.4.3 Summary and implications for future work

Analyses of time series data using quasi-Poisson regression methods showed small but significant increases in the proportion of children <2 years of age admitted to Patan Hospital with fast breathing pneumonia and chest indrawing pneumonia in comparison to non-pneumonia admissions, temporally associated with the introduction of 10-valent PCV to the infant immunisation schedule. In these analyses there was no effect of 10-valent

PCV on NP carriage of vaccine-type pneumococci. The limitations to the data, particularly in regard to the hypothesised large and pervasive effects as a result of earthquakes and political unrest in Nepal immediately prior to and during the introduction of 10-valent PCV are discussed in detail above.

As an alternative, I used the indirect cohort method that partly avoids confounding by temporal associations. This showed null vaccine effectiveness against outcomes of clinical pneumonia and pneumonia with endpoint consolidation that are non-specific for pneumococcal pneumonia in this age group (see Chapter 3), but high effectiveness against NP carriage of vaccine-type pneumococci. Given that NP carriage of pneumococci is hypothesised to be a prerequisite for pneumococcal pneumonia, one might expect that in a larger dataset vaccine effectiveness against non-specific endpoints might be detectable. An important caveat is that the inclusion of children who had received a single dose of 10-valent PCV within the "unvaccinated" group is likely to bias results towards a null effect of the vaccine.

In addition, at Patan Hospital, I have showed that the burden of invasive pneumococcal disease is predominantly in children >2 years of age (see Chapter 3). If the age distribution of invasive pneumococcal disease is indicative of that of pneumococcal pneumonia, then vaccine effectiveness against pneumonia will be detectable with increased power once the birth cohort of children eligible for 10-valent PCV reaches this age group. Further data are needed.

5

ALS for the detection of pneumococcal responses from antibody-secreting cells

5.1 Introduction

Current diagnostic tests for the aetiology of childhood pneumonia are either insensitive to the causative organism, or non-specific (see Chapter 1). An alternative diagnostic algorithm would examine the childhood immune response to the pathogen. Recently activated antibody-secreting cells (ASCs or plasmablasts) are known to circulate in the blood of patients with respiratory syncytial virus (RSV) infection[160], dengue virus infection[162], bacteraemia with Gram-negative organisms[261] and bacteraemic pneumococcal pneumonia[153], amongst other infections, between days 4 and 7 of illness. These cells become undetectable in peripheral blood by one month following the infection. Similarly, ASCs have a predictable peak at day 7 following mucosal and intramuscular immunisation with a number of antigens[161, 262–264], before becoming

undetectable in the peripheral circulation. A proportion of ASCs are antigen-specific, depending upon the immune stimulus[121]. Examination of the ASC response to infection may therefore yield both antigen-specific and time-specific data that could be used for the aetiological diagnosis of acute infections[121].

The antibodies from lymphocyte supernatant (ALS) assay was developed to study the secretions of ASCs from adults and children with active infection[138], and is detailed in Section 2.4.1. In brief, peripheral blood mononuclear cells are separated from a fresh whole blood sample from patients with infection, suspended in cell culture medium, and incubated (typically) for 48 hours. Following incubation the cells and supernatant are separated with a centrifugation step, with the supernatant presumed to be enriched for antibodies actively secreted by ASCs within the suspension. Assay of ALS may provide sensitive and specific diagnostics, with the practical advantage that it can be easily stored, transported, and analysed by ELISA or other methods in batches away from a routine clinical laboratory[121]. Assay of ALS has been pioneered as a diagnostic for presumed enteric fever[145, 146], tuberculosis[149, 150, 265], and to study vaccine responses[141], but theoretically could be used on any infection that elicits an ASC response. See Carter et al. (2017)[121] and Table 1.3.2 for further details.

Pneumococci are the most prevalent organism isolated from the blood of children with pneumonia admitted to Patan Hospital in Kathmandu, Nepal (see Chapter 3)[4, 68]. However, data from vaccine probe studies in other settings suggest that the proportion of pneumonia attributable to pneumococcal infection is considerably higher than that described by blood culture[3, 33]. The Kathmandu Valley is therefore an important epidemiological context in which to assess the development of ALS-based methods for the accurate diagnosis of pneumococcal pneumonia.

I aimed to assess the utility of assay of ALS for the diagnosis of pneumococcal pneumonia in children admitted to Patan Hospital with a clinical diagnosis of pneumonia enrolled to the PANDIT study. Almost all patients were concurrently enrolled into the Inpatient Pneumonia Study detailed in Chapters 2 and 3. I hypothesised that assay of ALS

to pneumococcal capsular polysaccharides (PS) may be a useful serotype-specific diagnostic test for pneumococcal pneumonia in these children.

Given that young children are known to generate poor immune responses against capsular polysaccharides unless presented with protein antigens[137, 266, 267] I hypothesised that assay of ALS to pneumococcal proteins may provide an alternative (although non-serotype-specific) diagnostic test for pneumococcal pneumonia in these children. For this analysis a FMIA was used to detect IgG ALS to 5 pneumococcal proteins that were selected as those widely used within the assessment of serological responses to pneumococcal infection[95] and carriage[100]. These were serine threonine kinase protein C (Stkp-C), pneumolysin (Ply), choline binding protein A (CbpA), secreted 45 kDa protein (PcsB), and pneumococcal histidine triad D (PhtD). This FMIA was undertaken at the Finnish National Institute for Health and Welfare (THL, Helsinki) using previously published methods[268]. Here I used all ALS samples taken from 348 children at enrolment to the study (within 48 hours of admission to hospital), and a further 20 samples from children from the community ("healthy controls").

In the absence of a "gold standard" test against which to test novel diagnostic tests for pneumococcal pneumonia I developed a series of "pneumonia aetiology comparator standards" based on proxy markers of pneumococcal pneumonia in children enrolled to the Inpatient Pneumonia Study (see Chapter 3). These comparator standards were identical for assay of ALS to pneumococcal capsular polysaccharides, and to pneumococcal proteins.

5.2 Methods

Ethics statement

Parents of all children enrolled to the study gave witnessed, written, informed consent before enrolment of their child into the study. The Nepal Health Research Council (reference 286/2014) and the Oxford Tropical Research Ethics Committee (reference 09/15) approved the study protocols.

Study site and population

Details of the study site are described in Section 2.1.

Summary of clinical, radiological, laboratory and microbiological methods

The background to studies of childhood pneumonia at Patan Hospital is described in Section 2.3.

5.2.1 Enrolment and sampling

The investigation of ALS for the detection of pneumococcal responses commenced enrolment on 26th February 2015 and concluded enrolment on 30th September 2016. Children $\geq$2 months (60 days) of age and up to 14 years of age who were being admitted to Patan Hospital with a clinical diagnosis of pneumonia were eligible for the PANDIT study. From September 2015 children $\geq$2 months (60 days) of age and up to 14 years of age who were diagnosed with clinical pneumonia at Patan Hospital, but not admitted to hospital, were also deemed eligible for enrolment.

The enrolment process, and non-specific laboratory methods for the PANDIT study are detailed in Section 2.3. In brief, the clinical diagnosis of pneumonia was made by the admitting paediatric team. Blood results and chest radiograph results were typically not available at the point of diagnosis and decision to admit to hospital. The decision to admit to hospital was based on the *IMCI* guidelines[94] with a degree of clinical discretion. All admissions were prospectively reviewed for eligibility for the PANDIT study. A chest radiograph and a single nasopharyngeal (NP) swab for culture of pneumococci was taken on all children with pneumonia enrolled into the PANDIT study, except a small number of children treated as outpatients on whom a chest radiograph was not done. Blood was taken aseptically from all enrolled children with pneumonia on admission for blood culture, full blood count and white cell differential in Patan Hospital Laboratory. Following storage at -80°C and shipment to the UK, plasma CRP concentration was retrospectively measured at the Biochemistry Laboratory at Oxford University Hospitals NHS Foundation Trust. Blood for culture was inoculated into a Bactec Peds Plus culture bottle (Becton

Dickinson, USA) and incubated aerobically using automated Bactec machines at Patan Hospital.

In addition 50 child healthy controls were enrolled into the PANDIT study to provide an additional comparator group for the analysis of ALS for the detection of pneumococcal responses from ASCs. A blood sample and a single NP swab for culture of pneumococci was taken from these children. Healthy controls were all local children,10 months of age, who had received three doses of 10-valent PCV, with at least four weeks between the third dose of PCV (or any other vaccine) and blood sampling. Full details of radiological and microbiological methods are in Section 2.3.

Samples for assay of ALS

An additional 3 ml of blood was taken from children enrolled to the PANDIT study. This was either taken at the same venepuncture as sampling for blood culture, full blood count and white cell differential, or within 48 hours of admission. 2 ml of blood was immediately placed into a heparinised 15 ml centrifuge tube, transported to the Microbiology Laboratory and processed (see below and Section 2.4.1). The remaining 1 ml of blood was immediately placed into RNA stabilising reagents in a purpose-designed collection tube (Tempus Blood RNA Tube, Applied Biosystems, USA) and immediately stored at -80°C (see 2.5.1). All children with pneumonia enrolled in the study were offered the opportunity to return to Patan Hospital to donate a further sample of 2 ml of blood at 6–8 weeks following their initial admission. Blood from healthy control children was processed as detailed above.

5.2.2 Development of comparator standards for pneumonia aetiology

I developed a series of pneumonia aetiology comparator standards prior to analysis of results of assays of ALS (or transcriptomic analyses, see Chapter 6). The comparator standards were hypothesised to provide a trend from a high proportion of children with pneumococcal pneumonia to a low proportion of children with pneumococcal pneumonia (see Figure 5.1). For clarity, I italicise these comparator standards in this chapter.

For these pneumonia aetiology comparator standards, children from whom a bacterial pathogen was cultured from blood or pleural fluid were classified as "*definite pneumococcal*" pneumonia or "*definite other bacterial*" pneumonia.

Children from whom no bacterial pathogen was cultured but with a plasma CRP concentration of $\geq$60 mg/l and NP carriage of serotype 1 pneumococci were classified as "*probable pneumococcal*" pneumonia. (Serotype 1 pneumococci are the major cause of childhood invasive pneumococcal disease, and NP carriage of serotype 1 pneumococci highly associated with pneumonia in children at Patan Hospital; Figure 3.9.)

Children from whom no bacterial pathogen was cultured, with a CRP concentration of $\geq$60 mg/l, but no NP carriage of serotype 1 pneumococci were classified as "*probable bacterial*" pneumonia. A CRP threshold of $\geq$60 mg/l was chosen on the basis of previously published literature [111, 112] and review of the CRP concentration from children with invasive pneumococcal disease in the Inpatient Pneumonia Study (Chapter 3).

Children from whom no bacterial pathogen was cultured with a CRP concentration <60 mg/l, a chest radiograph that showed either infiltrates or was normal and with a neutrophil count <12 x 10^3/mm^3 were classified as "*probable viral*" pneumonia. No viral testing was done on children within this study.

Children with a CRP concentration <60 mg/l but with a missing chest radiograph, or radiograph with endpoint consolidation or neutrophil count $\geq$12 x 10^3/mm^3 were classified as "*unknown*". In addition a group of 50 healthy children were recruited as controls as described above.

The comparator standards, and the number of children enrolled into each comparator standard are described in Figure 5.1. For convenience *definite other bacterial* pneumonia is shown as a category next to *definite pneumococcal* pneumonia. However, I hypothesised that children with *definite other bacterial* pneumonia did not have pneumococcal (co-)infection, and this comparator standard is depicted adjacent to *probable viral* pneumonia in the Results section.

5.2.3 ALS methods

Methods for the generation of ALS are detailed in Chapter 2. In brief, 2 ml of fresh whole blood was separated into plasma, peripheral blood mononuclear cells (PBMCs) and a mixture of red cells and polymorphonuclear cells within 4 hours of sampling by centrifugation over Ficoll media at 1.077 g/ml (Histopaque 1077, Sigma-Aldrich) at 400 g for 20 min with minimal acceleration and deceleration. The PBMC layer was separated manually with a sterile Pasteur pipette and washed by resuspension into R0 medium and centrifugation. All samples for which data is presented in this Chapter underwent a second wash step to ensure removal of contaminating serum from the PBMCs. Following washing and manual counting and calculation of the number of PBMCs in the sample, PBMCs were resuspended in R10 medium at 1 x 10^7 PBMCs/ml, placed into a cell culture plate and incubated at 35–37°C for 48 hours. Following incubation, the cell suspension was separated by centrifugation, with the supernatant (ALS) stored at -80°C at Patan Hospital with the addition of a small volume of 40X protease inhibitors (25 μl per 1 ml of ALS). Samples were shipped to Oxford Vaccine Group on dry ice in two batches. Aliquots were made in a single defrosting step, and distributed on dry ice to relevant laboratories. Samples were prioritised for use in the fluorescent multiplexed bead-based immunoassay (FMIA) to detect IgG to pneumococcal proteins, following examination of pilot data from the FMIA to pneumococcal capsular polysaccharides (both detailed below).

5.2.4 FMIA to 13 pneumococcal capsular polysaccharides

For this analysis a FMIA was used to detect IgG ALS to capsular polysaccharides contained within the 13-valent PCV (PS1, PS3, PS4, PS5, PS6A, PS6B, PS7F, PS9V, PS14, PS18C, PS19A, PS19F and PS23F). FMIA to detect ALS to pneumococcal capsular polysaccharides was done on the acute ALS samples from 100 children. These children were enrolled between November 2015 and June 2016 (ALS was generated from samples from the period February 2015 to October 2015 with a slightly different method, data not shown). All available acute samples from children with *definite pneumococcal* pneumonia, *probable pneumococcal* pneumonia and *definite other bacterial* pneumonia

were used. Samples from other comparator standards were selected using computer-generated randomisation from patients classified for each of the other comparator standards until the predefined number of samples was selected.

The FMIA of IgG ALS to pneumococcal capsular polysaccharides was undertaken at the Dutch National Institute for Public Health and the Environment (RIVM, Bilthoven) using previously published methods[269] that have been used to assess pneumococcal capsular polysaccharide antibody seroprevalence[270]. FMIA of IgA ALS to pneumococcal capsular polysaccharides was undertaken in a small number of samples and showed limited utility (data not shown).

The background to methods for the FMIA of IgG to pneumococcal capsular polysaccharides are detailed in Elberse et al. (2010) and Pickering et al. (2002)[269, 271]. In brief, samples of ALS were diluted and measured at two dilutions, 1/25 and 1/100, with the mean concentration of the assay was reported as the final result. ALS was diluted in adsorbent buffer containing 15 μg/ml of cell wall polysaccharide (CWPS Multi, Statens Serum Institut, Denmark) and 5% antibody-depleted human serum (ADHS, Valley Biomedical, USA) in phosphate buffer saline (PBS) and incubated at room temperature for 1 hour. The FMIA was performed as a 13-plex assay to pneumococcal capsular polysaccharides conjugated to carboxylated beads (Bio-Rad, USA) with the addition of beads carboxylated to pneumococcal cell wall polysaccharide (CWPS Multi, Statens Serum Institut, Denmark). A total of 3000 beads for each serotype-specific set were used per well. IgG antibodies in ALS were detected using RPE-conjugated goat anti-human IgG (Jackson ImmunoResearch, USA). A 10% weight/volume solution of lyophilised IVIG (Sanquin, Netherlands) was used as a reference serum (purified IgG from a pool of at least 1000 blood donors from the Dutch population). Data was acquired using BioPlex Software (Bio-Rad). IgG antibody concentrations were calculated by interpolating median fluorescence intensity to standard curve using a five parameter logistic regression line. The lower limits of quantitation for each antigen were: PS1 = 7 ng/ml, PS3 = 9 ng/ml, PS4 = 2 ng/ml, PS5 = 4 ng/ml, PS6A = 11 ng/ml, PS6B = 8 ng/ml, PS7F = 9 ng/ml, PS9V = 6 ng/ml, PS14 = 18 ng/ml, PS18C = 8 ng/ml, PS19A 11 ng/ml, PS19F = 36 ng/ml, PS

23F = 9 ng/ml, with measurements from samples below the limit of quantitation set at 1 ng/ml below the limit of quantitation.

5.2.5 FMIA to 5 pneumococcal proteins

FMIA to detect ALS to pneumococcal proteins was done on the acute ALS samples from 348 children (all available samples), regardless of comparator standard. In addition FMIA to detect ALS to pneumococcal proteins was done on 20 randomly chosen samples from healthy control children.

The full methods of the FMIA of IgG to pneumococcal proteins are detailed in Andrade et al. (2016)[106]. In brief, samples of ALS were diluted to 1/25 in deionised water. The FMIA was performed as a 5-plex assay with 1200 beads per region per well were used. IgG antibodies in ALS were detected using RPE-conjugated goat anti-human IgG (Jackson ImmunoResearch). Pneumococcal reference standard serum 007sp (NIBSC, UK) was used as the reference with an arbitrary assigned concentration of 1000 units/ml for each anti-pneumococcal antibody. Data was acquired using the xPONENT software v.4.2 (Luminex Corp, USA). The IgG antibody concentration of test samples was calculated by interpolating median fluorescence intensity to a seven point dilution standard curve using an asymmetrical five parameter logistic regression line. Two control serum samples were included in each plate to assess the assay-to-assay variation. The lower limits of quantitation for each antigen were: CbpA = 0.02 units/ml, PcsB = 0.03 units/ml, PhtD = 0.02 units/ml, Ply = 0.11 units/ml, StkP-C = 0.04 units/ml, with measurements from samples below the limit of quantitation set at 0.01 units/ml below the limit of quantitation. IgA measurements were made on a small number of samples – these pilot data showed low concentrations of ALS (data not shown). FMIA of IgA ALS to pneumococcal proteins was undertaken in a small number of samples and showed limited utility (data not shown).

Receiver operating characteristics (ROC) curves were created for concentrations of ALS to each pneumococcal protein antigen between children with *definite pneumococcal* pneumonia (i.e. blood or pleural fluid culture positive cases) and children with *definite*

other bacterial pneumonia (i.e. blood culture positive for other bacterial pathogens). The R package $OptimalCutpoints$[272] was used to generate an optimal threshold of ALS concentration against each pneumococcal protein antigen by minimising the probability of a false positive or false negative result (Youden's index). Children with *definite other bacterial* pneumonia were chosen as the comparator group, since they were likely to share many of the characteristics of children with *definite pneumococcal* pneumonia (e.g. an active bacterial infection), but these children were very unlikely to have an unidentified co-infection with *S. pneumoniae*. They therefore provided a robust comparator group.

5.2.6 Assays to control antigens and further investigations

Specificity of FMIA to pneumococcal capsular polysaccharide "specific" IgG ALS

The specificity of FMIA of IgG to specific pneumococcal capsular polysaccharides was investigated by the addition of specific capsular polysaccharide before repeated measurement. Following initial quantitation of IgG to PS1 and PS3 using FMIA, 50 μg/ml of PS1 and PS3 (Statens Serum Institut) were added to 12 samples each of ALS and allowed to adsorb for 1 hour. IgG to capsular PS1 and capsular PS3 was then re-quantified in the "spiked" samples. These experiments were done at RIVM using the FMIA detailed above.

ELISA for the quantitation of pneumococcal specific IgG ALS

IgG to 9 specific pneumococcal polysaccharides was quantified in 10 samples of ALS using the WHO-validated ELISA for the quantitation of *S. pneumoniae* serotype-specific IgG (Pn PS ELISA)[273]. These assays were done at the WHO Pneumococcal Serology Reference Laboratory, UCL, London, UK.

ELISA for the quantitation of H. influenzae type B specific IgG ALS

The ability of ALS to detect IgG to other polysaccharide antigens was assessed using an ELISA to *H. influenzae* type b (Hib). Hib specific IgG ELISAs were done at RIVM using methods detailed by Mariani et al. (1998)[274].

5.3 Results

5.3.1 Enrolment and clinical classifications

During the study period, 610 children aged 2 months to 14 years were admitted to Patan Hospital with a clinical diagnosis of pneumonia. Of these, 514 (84%) were enrolled into the study. An additional 40 children aged 2 months to 14 years with a clinical diagnosis of pneumonia made at Patan Hospital, but who were not admitted, were enrolled to the study, totalling 554 children. An additional group of 50 child healthy controls were enrolled.

By pneumonia aetiology comparator standard, 11 children were classified as *definite pneumococcal* pneumonia, 11 children as *definite other bacterial* pneumonia, 5 children as *probable pneumococcal* pneumonia, 98 children as *probable bacterial* pneumonia, 199 children as *unknown* pneumonia and 230 children as *probable viral* pneumonia. Figure 5.1 describes the numbers of children enrolled to each comparator standard, and the sample availability for these children.

Sample availability was determined by the volume of blood that was taken from children enrolled into the study (with an absolute requirement of 1 ml of blood for generation of ALS), and the methods used to wash PBMCs. As described in Section 2.4.1 samples from the first 150 children enrolled in the study (including 4 cases of *definite pneumococcal* pneumonia) were from PBMCs that were washed only once to remove any potentially contaminating serum. Samples from these first 150 children are not reported.

5.3.2 FMIA for the quantitation pneumococcal capsular polysaccharide specific IgG ALS

In total, acute ALS samples from 100 children were assayed for IgG to pneumococcal capsular polysaccharides. Samples were tested from 5 children with *definite pneumococcal* pneumonia, 3 children with *probable pneumococcal* pneumonia, 39 children with *probable bacterial* pneumonia, 39 children with *unknown* aetiology, 11 children with

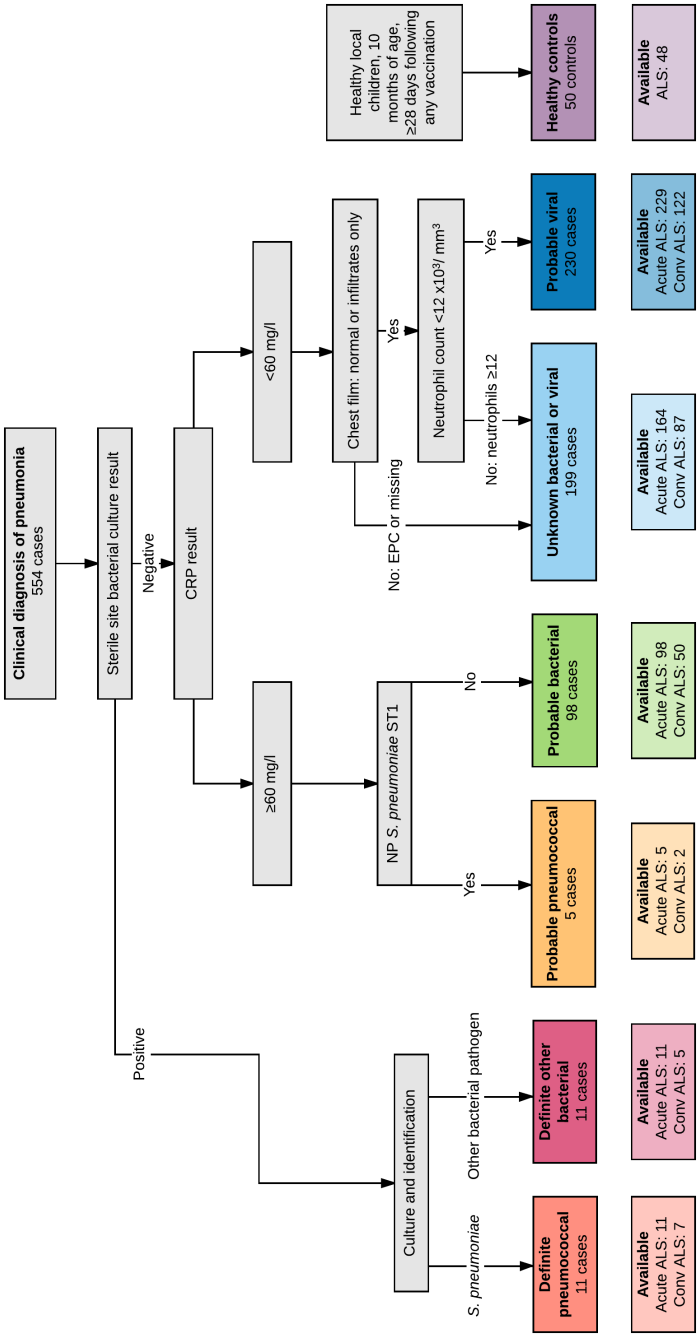


Figure 5.1: Enrolment and sample availability for the study of ALS as a diagnostic test for the aetiology of pneumonia (within the PANDIT study). Abbreviations: ST1, serotype 1; conv, convalescent.

probable viral pneumonia and 3 children with *definite other bacterial* pneumonia. No samples from healthy controls were tested.

All children with *definite pneumococcal* pneumonia were bacteraemic for *S. pneumoniae*. Children with *definite pneumococcal* pneumonia had a lower median age (5.5 years, IQR 4.3, 9.1) in comparison to children with *definite other bacterial* pneumonia (8.9 years, IQR 4.8, 8.9; Mann-Whitney U-test, p = 0.006). Children with *probable pneumococcal* pneumonia, *probable bacterial* pneumonia, *unknown*, and *probable viral* pneumonia had median ages of 6.8 years (IQR 4.8, 7.5), 3.1 years (IQR 1.9, 4.7), 0.9 years (0.5, 1.5) and 1.1 years (0.7, 1.4), respectively. By parental report, children with *definite pneumococcal* pneumonia had illness for a median of 4 days (IQR 4.0, 5.0) in comparison to a median of 7 days (IQR 4, 7) for children with *definite other bacterial* pneumonia (Mann-Whitney U-test, p < 0.001). Children with *probable pneumococcal* pneumonia, *probable bacterial* pneumonia, *unknown*, and *probable viral* pneumonia had illness for a median of 5 days (IQR 4.5, 6), 4 days (IQR 3, 7), 6 days (5, 8) and 6 days (3.5, 6), respectively. Additional data on plasma CRP concentrations, blood neutrophil count, the proportion of children with radiographic endpoint consolidation and NP carriage of pneumococcal serotypes are detailed in Table 5.1.

IgG ALS were non-specific to pneumococcal capsular polysaccharides

Concentrations of IgG ALS to pneumococccal capsular polysaccharides in 13-valent PCV ranged from 4.5 μg/ml (to PS3 in a child with *probable bacterial* pneumonia) to below the limit of quantitation. 21 children had IgG ALS below the level of quantitation to all capsular polysaccharides. Results for samples from all children on whom IgG ALS to pneumococcal capsular polysaccharides was quantified in the FMIA are shown in Figure 5.2, without classification into comparator standards. IgG ALS to PS3 (denoted as serotype) was greater than IgG to any other polysaccharide.

Visualisation of results was facilitated by display on a $\log_{10}$ scale, with samples ordered by concentration of IgG to PS3 (Figure 5.3). This shows a high degree of correlation between IgG ALS to pneumococcal capsular polysaccharides: if a high concentration of IgG ALS to PS3 is detected then high concentrations of IgG ALS to other polysaccharides are also

Table 5.1: Characteristics of children on whom IgG ALS to pneumococcal capsular polysaccharides was quantified by pneumonia aetiology comparator standard. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control.

	Comparator standard								
	DP	**PP**	**PB**	**U**	**PV**	**DOB**	**HC**	**Any pneumonia**	**p[a]**
n	5	3	39	39	11	3	0	100	
Male sex	4 (80%)	3 (100%)	19 (49%)	19 (49%)	4 (36%)	3 (100%)	-	52 (52%)	1
Age in years (median, IQR)	5.5 (4.3, 9.1)	6.8 (4.8, 7.5)	3.1 (1.9, 4.7)	0.9 (0.5, 1.5)	1.1 (0.7, 1.4)	8.9 (4.8, 8.9)	-	1.8 (0.8, 3.9)	0.006
2 months to <2 years	0	0	10 (28%)	29 (81%)	10 (91%)	1 (33%)	-	51 (51%)	
2 years to <5 years	2 (40%)	1 (33%)	17 (47%)	6 (47%)	1 (9%)	0	-	31 (31%)	
5 years to 14 years	3 (60%)	2 (67%)	9 (25%)	1 (2.8%)	0	2 (67%)	-	18 (18%)	
Length of illness (median, IQR)	4.0 (4.0, 5.0)	5.0 (4.5, 6.0)	4.0 (3.0, 7.0)	6.0 (5.0, 8.0)	6.0 (3.5, 6.0)	7.0 (4.5, 8.0)	-	6.0 (4.0, 7.0)	<0.001
CRP mg/l (median, IQR)	181 (102, 182)	107 (86, 157)	138 (88, 197)	12 (3.7, 26.8)	13 (5.9, 20)	90 (52, 138)	-	56 (13, 131)	
CRP mg/l $\geq$60 mg/l	4 (80%)	3 (100%)	36 (100%)	0	0	2 (67%)	-	48 (48%)	
Neutrophil count x 10^3/mm^3 (median, IQR)	25 (19, 27)	13 (12, 21)	13 (9, 16)	6.0 (3.8, 9.7)	7.2 (3.6, 8.4)	14 (12, 15)	-	9.3 (5.7, 14)	
Endpoint consolidation	5 (100%)	2 (67%)	33 (85%)	38 (97%)	0	2 (67%)	-	80 (80%)	
Pneumococcal carriage	1 (20%)	3 (100%)	15 (38%)	11 (28%)	10 (91%)	0 (	-	40 (40%)	
Pneumococcal carriage (no antibiotics)	1/2 (50%)	2/2 (100%)	7/21 (33%)	4/16 (25%)	8/9 (89%)	0	-	22/50 (44%)	

[a]p values calculated using χ^2 test without continuity corrections for two groups (for sex), or Mann-Whitney U-test (for age and length of illness prior to sampling) for a comparison between patients with *definite pneumococcal* and definite bacterial pneumonia.

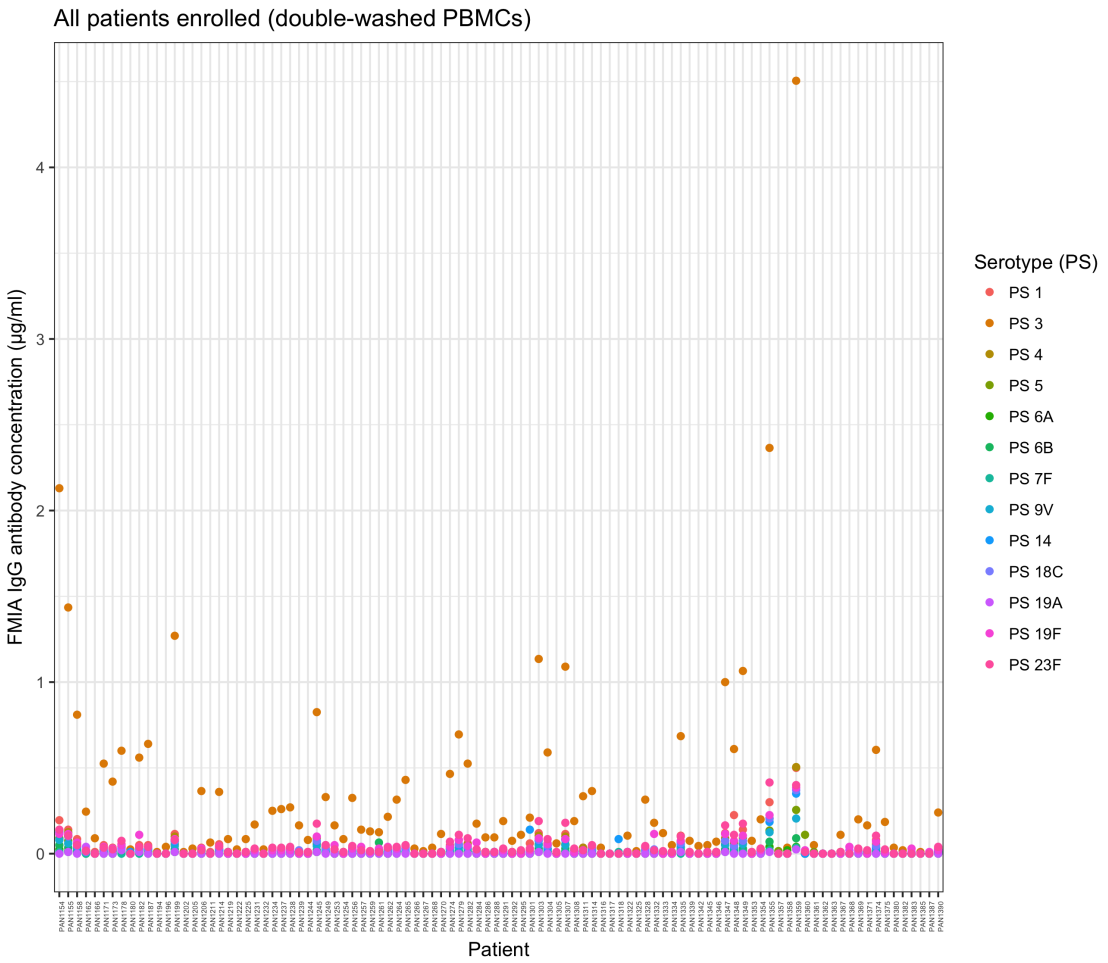


Figure 5.2: Quantification of IgG ALS to pneumococcal capsular polysaccharides in 100 children with pneumonia in Nepal by order of enrolment into the study. FMIA, fluorescent multiplexed bead-based immunoassay.

detected. The concentrations of IgG ALS to pneumococcal capsular polysaccharides did not correlate with NP carriage of pneumococci, regardless of whether pre-treatment with antibiotics was done (data not shown).

The finding of high IgG ALS to PS3, and the correlation between capsular polysaccharide concentrations more generally, was unexpected. To investigate this, purified PS1 or purified PS3 was added to aliquots of ALS from 12 children, in an effort to preadsorb specific antibodies to PS1 and PS3. Samples from these 12 children were chosen on the basis of availability (all were from children enrolled within the first 6 months of enrolment), and representation of a wide range of measured IgG ALS to PS1 and PS3 values. These "spiking experiments" did not reduce the measured concentrations of IgG ALS to PS1 or

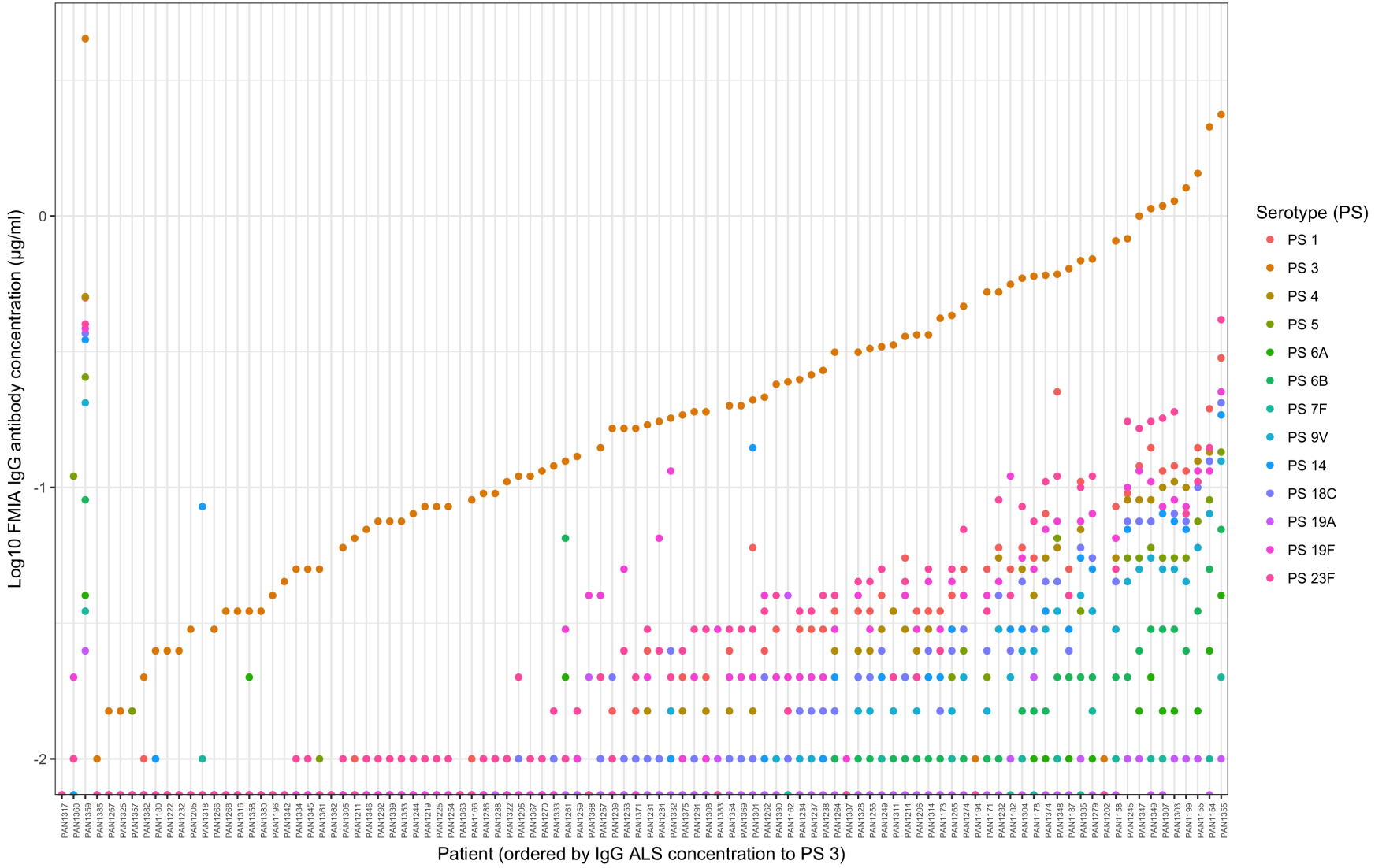


Figure 5.3: IgG ALS to pneumococcal capsular polysaccharides in 100 children (patients) with pneumonia in Nepal, ordered by levels of IgG to PS3 on the $\log_{10}$ scale, with NP carriage of pneumococcal serotype (if present) denoted. FMIA, fluorescent multiplexed bead-based immunoassay.

Table 5.2: Antibodies as measured by FMIA to purified capsular polysaccharides 1 and 3 following addition of ALS with 50 μg/ml of purified capsular polysaccharides (PS) of serotypes 1 and 3 ("spiking experiments").

	Capsular PS1			Capsular PS 3		
Participant	**Buffer**	**50 μg/ml PS1**	**50 μg/ml PS3**	**Buffer**	**50 μg/ml PS1**	**50 μg/ml PS3**
PAN1005	0.71	0.49	0.72	3.79	2.85	4.06
PAN1007	8.47	7.55	7.66	24.73	22.95	21.05
PAN1009	1.94	1.45	1.37	7.69	6.82	5.98
PAN1010	0.10	0.06	0.08	0.57	0.36	0.38
PAN1012	0.18	0.11	0.12	0.94	0.56	0.64
PAN1016	0.94	0.72	0.77	4.14	3.18	3.42
PAN1032	0.84	0.51	0.60	4.45	2.84	2.95
PAN1037	0.11	0.09	0.10	0.61	0.43	0.46
PAN1046	0.67	0.52	0.48	2.86	2.23	2.16
PAN1054	1.01	0.72	0.68	4.57	3.11	2.92
PAN1062	1.34	1.22	1.24	5.61	5.05	5.10
PAN1067	0.80	0.72	0.72	3.25	3.27	2.92

PS3 (Table 5.2). This suggested that the FMIA was detecting a signal other than that from PS1 or PS3 specific IgG ALS.

To test this hypothesis, aliquots of ALS from 10 children were tested using the WHO-validated ELISA for the quantitation of *S. pneumoniae* serotype-specific IgG (Pn PS ELISA)[273]. Again, samples were chosen on the basis of sample availability, but from across the entire period of enrolment, and include samples with a range of measured IgG ALS to capsular polysaccharides. Quantitation of IgG to 9 pneumococcal capsular polysaccharides by the Pn PS ELISA detected concentrations above the limit of quantitation in 2 of the 10 aliquots of ALS assayed (Figure 5.4). These two samples were from children enrolled in the pilot phase for the generation of ALS at Patan Hospital (not otherwise described here) with the possibility of plasma contaminating the ALS sample. The greatest concentration PS1 or PS3 specific IgG ALS was from child PAN1007, a boy of 8 years of age with no recent immunisations and with bacteraemia for *S. pneumoniae* serotype (ST) 1. The concentrations of IgG ALS to capsular polysaccharides measured by the Pn PS ELISA did not correlate with the concentrations of IgG ALS to capsular polysaccharides measured by FMIA (data not shown).

We also tested whether IgG ALS to non-pneumococcal vaccine polysaccharide and protein antigens were detectable in aliquots of ALS from these children with pneumonia. Concentrations of Hib specific IgG were between 0.42 μg/l and 3.99 μg/l in the 10 aliquots

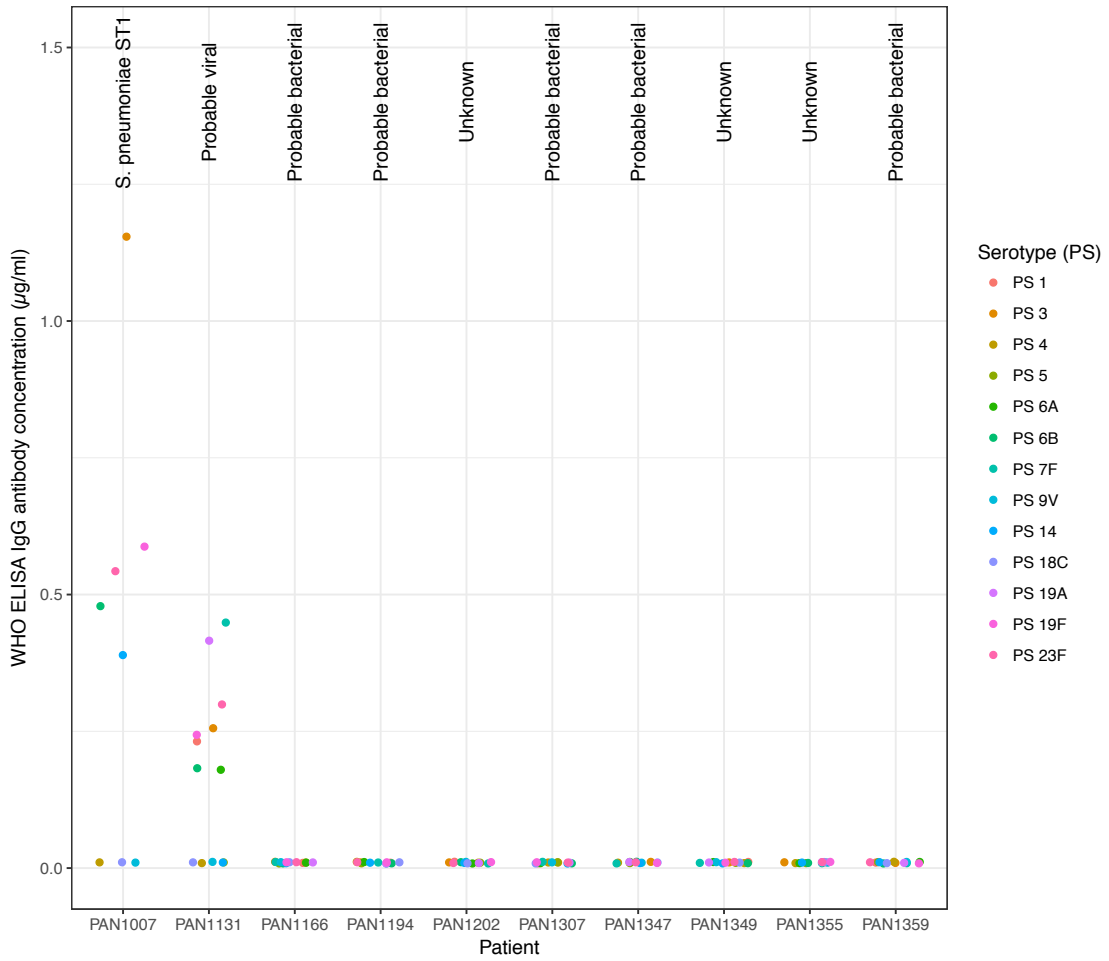


Figure 5.4: Antibodies to pneumococcal capsular polysaccharides in ALS from 10 children with pneumonia in Nepal (with classification of likely pneumonia aetiology, top).

of ALS assayed. The highest concentration of Hib specific IgG in the 12 aliquots of ALS assayed was from child PAN1007 (see above). Concentrations of IgG in the 12 aliquots of ALS assayed to protein antigens commonly used in vaccines (pertussis antigens: pertussis toxin, filamentous haemagglutinin antigen, pertactin; diphtheria toxoid and tetanus toxoid) are shown in Table 5.4. These are all low, and considerably below the threshold for the serological diagnosis of pertussis [275].

IgG ALS did not discriminate between pneumonia aetiology comparator standards

Concentrations of IgG ALS to pneumococcal capsular polysaccharides as measured by the FMIA are shown by comparator standard for each capsular polysaccharide in 13-valent PCV in Figure 5.5. There was no association between comparator standard and concentrations of capsular polysaccharide specific IgG ALS for any capsular polysac-

Table 5.3: Antibodies as measured by FMIA to *Haemophilus influenzae* (Hib) type B in ALS from patients with pneumonia. These children are from a population with high vaccine coverage against Hib.

	***Haemophilus influenzae* type b (μg/ml)**		
Participant	**Dilution 1:10**	**Dilution 1:100**	**Average**
PAN1005	0.32	0.53	0.42
PAN1007	1.31	6.67	3.99
PAN1009	0.68	1.40	1.04
PAN1032	0.90	1.04	0.97
PAN1046	0.32	0.71	0.51
PAN1047	0.61	1.02	0.81
PAN1048	0.31	0.73	0.52
PAN1053	0.53	1.13	0.83
PAN1054	0.48	1.30	0.89
PAN1064	0.72	2.52	1.62

Table 5.4: Antibodies as measured by FMIA to pertussis antigens, diphtheria toxin and tetanus toxoid in ALS. These children are from a population with high vaccine coverage for diphtheria-tetanus-pertusiss. PT, pertussis toxin; FHA, filamentous haemagglutinin antigen; Prn, pertactin.

	Pertussis			**Diphtheria toxoid (μg/ml)**	**Tetanus toxoid (μg/ml)**
Participant	**PT (μg/ml)**	**FHA (μg/ml)**	**Prn (μg/ml)**		
PAN1054	1.0	1.0	0.8	0.019	0.013
PAN1059	1.0	1.0	0.3	0.007	0.004
PAN1060	1.1	1.0	0.5	0.017	0.011
PAN1062	1.1	1.0	0.9	0.020	0.016
PAN1064	2.9	0.6	2.6	0.065	0.033
PAN1067	1.8	1.0	1.4	0.028	0.029
PAN1071	1.0	1.0	1.0	0.004	0.006
PAN1073	1.0	1.0	1.0	0.002	0.003
PAN1075	1.0	1.0	0.5	0.012	0.010
PAN1078	1.0	1.0	1.0	0.004	0.009
PAN1079	1.0	1.0	1.0	0.005	0.004

charide. Concentrations of IgG ALS to PS1 by comparator standard are shown in Figure 5.6, with the serotype of invasive pneumococcal disease denoted in text. In these data, there was no association between the serotype of invasive pneumococcal disease and the magnitude of IgG ALS to that corresponding capsular polysaccharide.

In summary, we detected IgG ALS to pneumococcal and Hib polysaccharides in these children with pneumonia, but the results did not appear to specific in children with definite pneumococcal pneumonia, and did not discriminate between comparator standards.

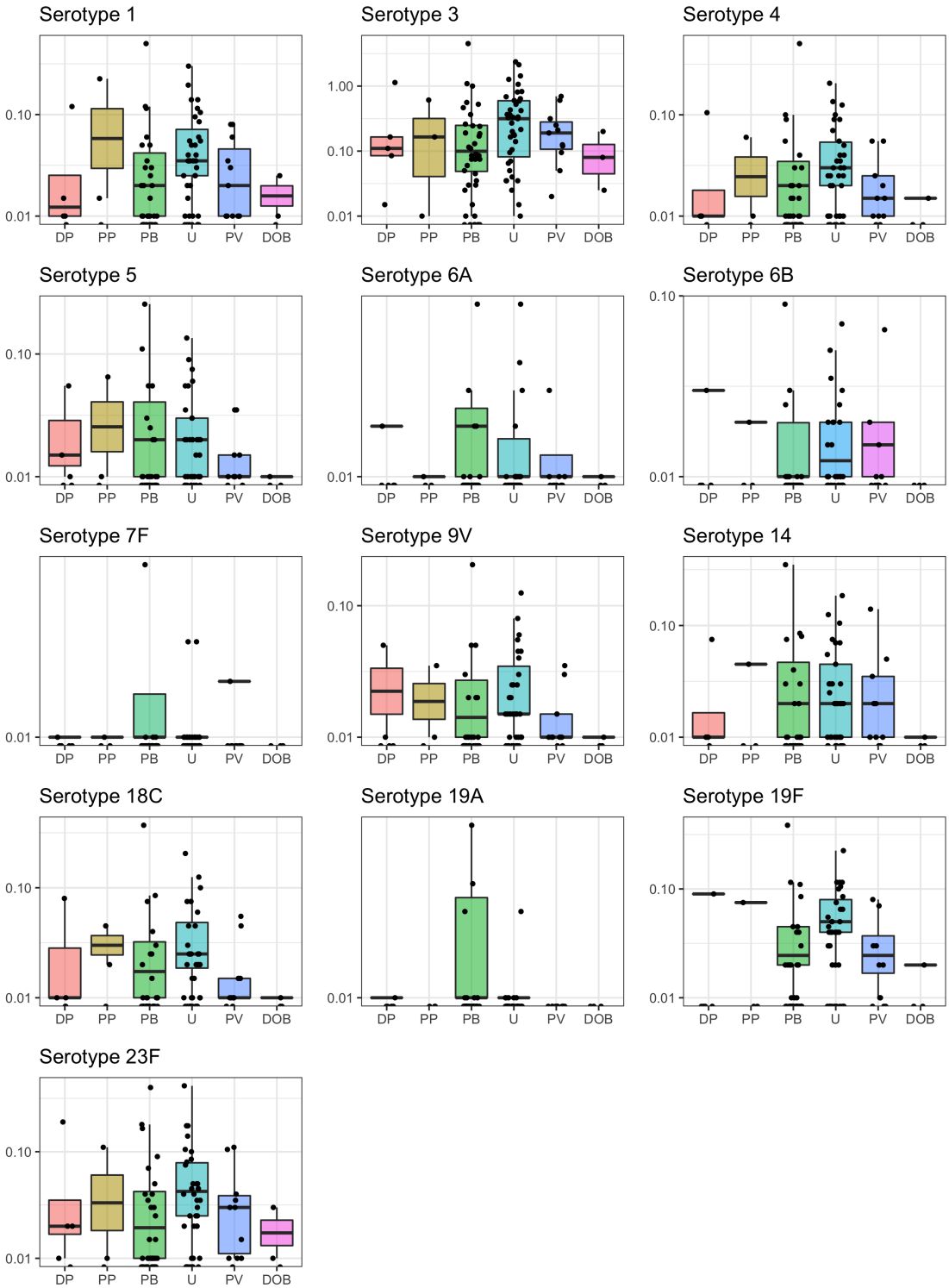


Figure 5.5: ALS to pneumococcal capsular polysaccharides (within 13-valent PCV) by pneumonia aetiology comparator standard. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*.

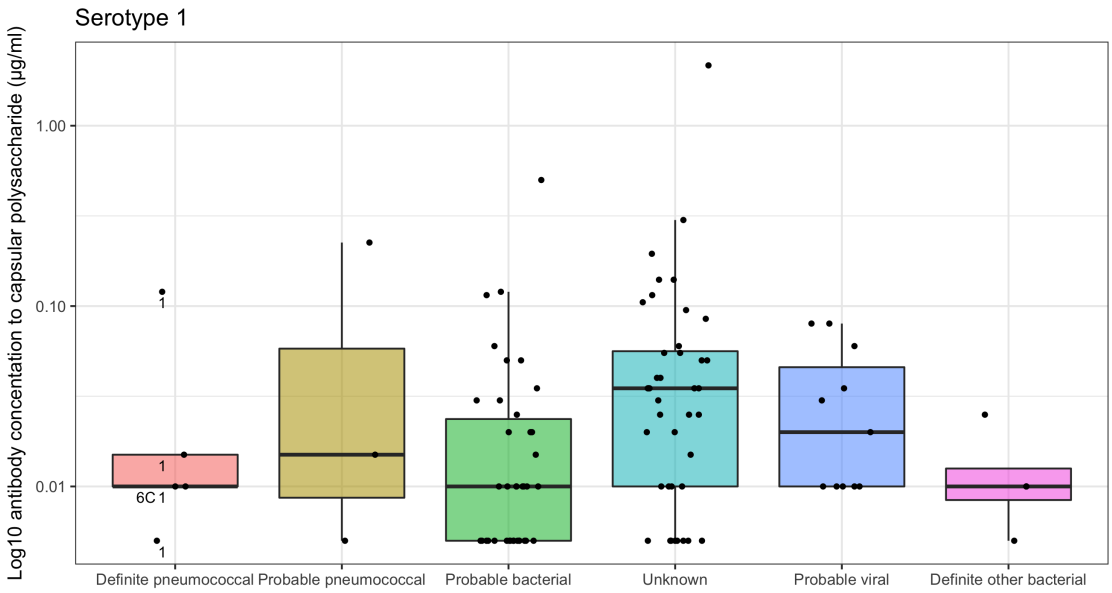


Figure 5.6: ALS to serotype 1 capsular polysaccharide by pneumonia aetiology comparator standard. Invasive serotypes are denoted for pneumococcal pneumonia.

5.3.3 Results of FMIA to detect ALS to pneumococcal proteins

Characteristics of children enrolled

Acute ALS samples from 348 children with pneumonia and 20 healthy controls were tested by FMIA to pneumococcal proteins. Samples were tested from 8 children with *definite pneumococcal* pneumonia, 4 children with *probable pneumococcal* pneumonia, 67 children with *probable bacterial* pneumonia, 107 children with *unknown* aetiology, 155 children with *probable viral* pneumonia and 7 children with *definite other bacterial* pneumonia. This consisted of acute samples from all children in the post-pilot period of data collection.

Of children with *definite pneumococcal* pneumonia, seven were bacteraemic for *S. pneumoniae* and one child had *S. pneumoniae* identified from pleural fluid alone. Amongst children with *definite other bacterial* pneumonia the following *a priori* defined pathogens were isolated from blood: *Acinetobacter* spp. (2 children), *E. coli* (1 child), *N. meningitidis* (1 child), *Pseudomonas* spp. (1 child), *S. aureus* (2 children). Children with *definite pneumococcal* pneumonia had a higher median age (5.6 years, IQR 4.2, 8.1) in comparison to children with *definite other bacterial* pneumonia (4.0 years, IQR 0.7, 8.2; Mann-Whitney

U-test, $p < 0.001$). Children with *probable pneumococcal* pneumonia, *probable bacterial* pneumonia, *unknown*, and *probable viral* pneumonia had median ages of 7.6 years (IQR 5.8, 8.3), 2.7 years (IQR 1.8, 5.0), 1.1 years (0.6, 2.4) and 1.0 years (0.5, 1.8), respectively. By parental report, children with *definite pneumococcal* pneumonia had illness for a median of 4 days (IQR 3.8, 5.0) in comparison to a median of 7 days (IQR 5.5, 8.5) for children with *definite other bacterial* pneumonia (Mann-Whitney U-test, $p < 0.001$). Children with *probable pneumococcal* pneumonia, *probable bacterial* pneumonia, *unknown*, and *probable viral* pneumonia had illness for a median of 4.5 days (IQR 3.3, 5.5), 5 days (IQR 3.5, 7), 6 days (4, 8) and 5 days (4, 7), respectively. Additional data on plasma CRP concentrations, blood neutrophil count, the proportion of children with radiographic endpoint consolidation and NP carriage of pneumococcal serotypes are detailed in Table 5.5.

Concentrations of IgG ALS to the 5 pneumococcal proteins (CbpA, PcsB, PhtD, Ply and Stkp-C) are shown in Table 5.6. Visual inspection of the distribution of concentrations of IgG ALS revealed a greater range of concentrations from children over 2 years of age, and in samples taken between 2 and 10 days of reported illness (data not shown), which was anticipated by a recent review of the topic[121].

Children with *definite pneumococcal* pneumonia ($n = 8$) had significantly higher concentrations of IgG ALS to CbpA, PcsB, PhtD and Stkp-C, but not Ply in comparison to healthy control children ($n = 20$, see Table 5.6). However, when compared against the smaller group of children with *definite other bacterial* pneumonia, the comparison was only significant for PhtD. For all IgG ALS to all 5 pneumococcal proteins there was a significant trend towards lower concentrations across the pneumonia aetiology comparator standards (Cuzick's test for trend, Table 5.6 and Figure 5.8).

Receiver operated characteristic (ROC) curves for IgG ALS to each pneumococcal protein antigen, and optimum thresholds to discriminate between *definite pneumococcal* and *definite other bacterial* pneumonia were derived, as shown in Figure 5.7 and Table 5.7. As anticipated in Table 5.6, IgG specific for CbpA and PhtD showed the greatest discriminatory ability, followed by PcsB. IgG ALS to Ply and Stkp-C were

Table 5.5: Characteristics of children on whom ALS to pneumococcal proteins was measured by pneumonia aetiology comparator standard. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control. The seven children with *definite other bacterial* pneumonia had the following pathogens isolated from blood: *Acinetobacter* spp. (two children), *E. coli* (one child), *N. meningitidis* (one child), *Pseudomonas* spp. (one child), *S. aureus* (two children).

	Comparator standard								
Protein	**DP**	**PP**	**PB**	**U**	**PV**	**DOB**	**HC**	**Any pneumonia**	**p**[a]
n	8	4	67	107	155	7	20	348	
Male sex	7 (88%)	4 (100%)	39 (58%)	60 (56%)	94 (61%)	4 (57%)	–	208 (60%)	0.2
Age in years (median, IQR)	5.6 (4.2, 8.1)	7.6 (5.8, 8.3)	2.7 (1.8, 5.0)	1.1 (0.6, 2.4)	1.0 (0.5, 1.8)	4.0 (0.7, 8.2)	–	1.4 (0.6, 2.7)	<0.001
2 months to <2 years	0	0	20 (30%)	69 (64%)	123 (79%)	4 (57%)	–	216 (62%)	
2 years to <5 years	3 (38%)	1 (25%)	29 (43%)	26 (24%)	29 (19%)	0	–	88 (25%)	
5 years to 14 years	5 (63%)	3 (75%)	18 (27%)	12 (11%)	3 (0.2%)	3 (43%)	–	44 (13%)	
Length of illness (median, IQR)	4 (3.8, 5.0)	4.5 (3.3, 5.5)	5.0 (3.5, 7.0)	5.0 (4.0, 8.0)	5.0 (4.0, 7.0)	7.0 (5.5, 8.5)	–	5.0 (4.0, 7.0)	<0.001
CRP mg/l (median, IQR)	141 (56, 183)	133 (97, 171)	115 (82, 183)	16 (4.0, 28)	7.8 (2.4, 17)	13 (4.2, 52)	–	16 (4.2, 53)	
CRP $\geq$60 mg/l	5 (63%)	4 (100%)	67 (100%)	0	0	2 (29%)	–	78 (22%)	
Neutrophil count x 10^3/mm^3 (median, IQR)	17 (10, 25)	20 (13, 28)	13 (8.3, 17)	8.5 (4.8, 15.0)	4.8 (3.5, 7.4)	9.7 (5.0, 12.3)	–	7.2 (4.0, 35.8)	
Endpoint consolidation	7 (88%)	3 (75%)	48 (72%)	56 (52%)	0	2 (29%)	–	116 (33%)	
Pneumococcal carriage	2 (25%)	4 (100%)	25 (37%)	30 (28%)	41 (26%)	1 (14%)	–	123 (35%)	
Pneumococcal carriage (no antibiotics)	1/1 (50%)	2/2 (100%)	20 (56%)	23 (37%)	31 (38%)	0	–	77/186 (41%)	

[a]p values calculated using χ^2 test of proportions for sex, or Mann-Whitney U-test for age and length of illness prior to sampling for a comparison between patients with *definite pneumococcal* and definite bacterial pneumonia.

Table 5.6: IgG ALS to the pneumococcal proteins CbpA, PcsB, PhtD, Ply and Stkp-C by pneumonia aetiology comparator standard. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control. p trend refers to Cuzick's test for trend across the comparator standards, ranked by *a priori* estimated probability of true pneumococcal pneumonia.

	Comparator standard									
	DP	**PP**	**PB**	**U**	**PV**	**DOB**	**HC**	**p (DP vs HC)**	**p (DP vs DOB)**	**p trend**
CbpA (median, IQR)	0.10 (0.07, 0.23)	0.24 (0.15, 0.32)	0.06 (0.02, 0.13)	0.01 (0.01, 0.06)	0.01 (0.01, 0.05)	0.02 (0.01, 0.09)	0.01 (0.01, 0.01)	<0.001	0.092	<0.001
PcsB (median, IQR)	0.08 (0.05, 0.39)	0.32 (0.08, 0.39)	0.04 (0.02, 0.16)	0.02 (0.02, 0.10)	0.02 (0.02, 0.05)	0.05 (0.02, 0.07)	0.02 (0.02, 0.02)	0.001	0.176	<0.001
PhtD (median, IQR)	0.05 (0.04, 0.17)	0.09 (0.07, 0.16)	0.03 (0.01, 0.10)	0.01 (0.01, 0.06)	0.01 (0.01, 0.03)	0.02 (0.01, 0.01)	0.01 (0.01, 0.01)	<0.001	0.022	<0.001
Ply (median, IQR)	0.24 (0.10, 0.53)	0.30 (0.18, 0.52)	0.10 (0.10, 0.31)	0.12 (0.10, 0.39)	0.10 (0.10, 0.25)	0.17 (0.10, 0.32)	0.10 (0.10, 0.22)	0.163	0.583	0.014
Stkp-C (median, IQR)	0.10 (0.05, 0.35)	0.05 (0.03, 0.11)	0.03 (0.03, 0.11)	0.04 (0.03, 0.11)	0.03 (0.03, 0.07)	0.06 (0.05, 0.08)	0.03 (0.03, 0.04)	0.003	0.38	0.001

Table 5.7: Optimal thresholds and diagnostic accuracy (95% CIs) measures for IgG ALS concentrations to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia for five pneumococcal proteins: CbpA, PcsB, PhtD, Ply and Stkp-C.

	Protein				
	CbpA	**PcsB**	**PhtD**	**Ply**	**Stkp-C**
Threshold	0.06 units/ml	0.08 units/ml	0.04 units/ml	0.5 units/ml	0.26 units/ml
Sensitivity	0.88 (0.47, 1.00)	0.625 (0.24, 0.91)	0.875 (0.47, 1.00)	0.375 (0.09, 0.76)	0.375 (0.09, 0.76)
Specificity	0.714 (0.29, 0.96)	0.714 (0.29, 0.96)	0.714 (0.29, 0.96)	1 (0.59, 1.00)	1 (0.59, 1.00)
False positives	2/7	2/7	2/7	0/7	0/7
False negative	1/8	3/8	1/8	5/8	5/8

unable to discriminate these two comparator standards. The ALS concentrations to each pneumococcal protein antigen are described by pneumonia aetiology comparator standard in Figure 5.8, with a dashed line on plots for CbpA, PcsB and PhtD representing the optimal threshold for these antigens (as $\log_{10}$ values).

In these data, IgG ALS concentrations to PhtD were significantly higher in children with *definite pneumococcal* vs *definite other bacterial* pneumonia, and had identical or better diagnostic accuracy than other pneumococcal protein antigens to discriminate between *definite pneumococcal* and *definite other bacterial* pneumonia. Using the threshold derived from ROC curve analysis, I explored the association between the proportion of children with a "positive" result of IgG ALS to PhtD to the proportions of children with positive results to the proxy markers of pneumococcal pneumonia (radiographic endpoint consolidation and plasma CRP concentration $\geq$60 mg/l).

Of 348 enrolled children (excluding health controls), 116 (33%) had endpoint consolidation on chest radiograph. Of these 116 children, 43 (37%) had a positive result of IgG ALS to PhtD, in comparison to 38/232 (16%) of children without endpoint consolidation. The association between a positive result of IgG ALS to PhtD and endpoint consolidation was significant in univariate logistic regression, and when adjusted for age (odds ratio 2.11, 95% CI 1.20, 3.71; p = 0.009). IgG ALS to PhtD was correlated with plasma CRP concentrations in unadjusted analyses across all age groups (Pearson correlation, p = 0.03), but not when adjusted for age in linear regression models (p = 0.36). Of 348 enrolled children with pneumonia, 78 (22%) had a CRP concentration of $\geq$60 mg/l. Of these 78 children, 29 (37%) had a "positive" result of IgG ALS to PhtD, in comparison to

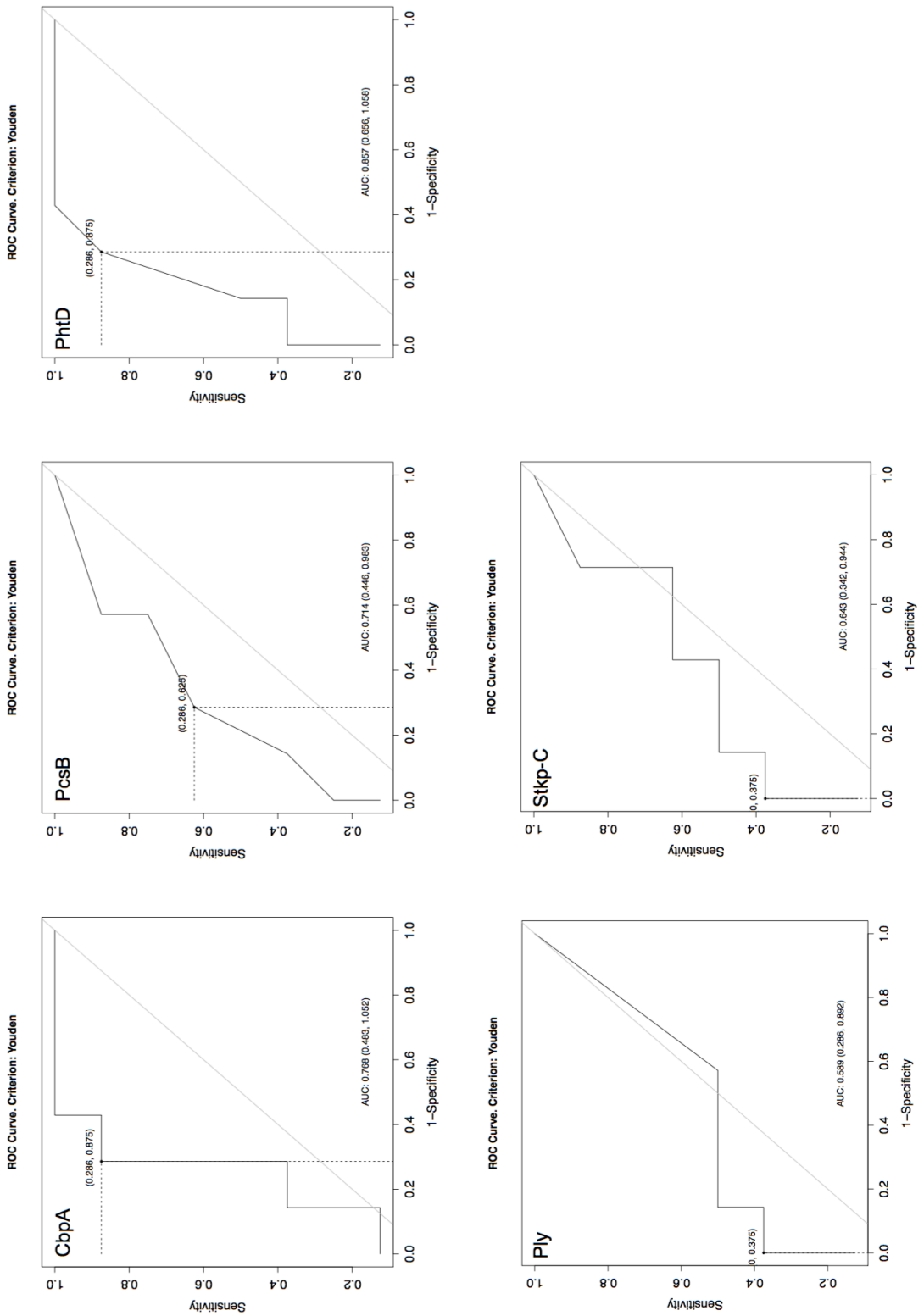


Figure 5.7: ROC curves to determine optimal thresholds for IgG ALS concentrations to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia for five pneumococcal proteins: CbpA, PcsB, PhtD, Ply and Stkp-C.

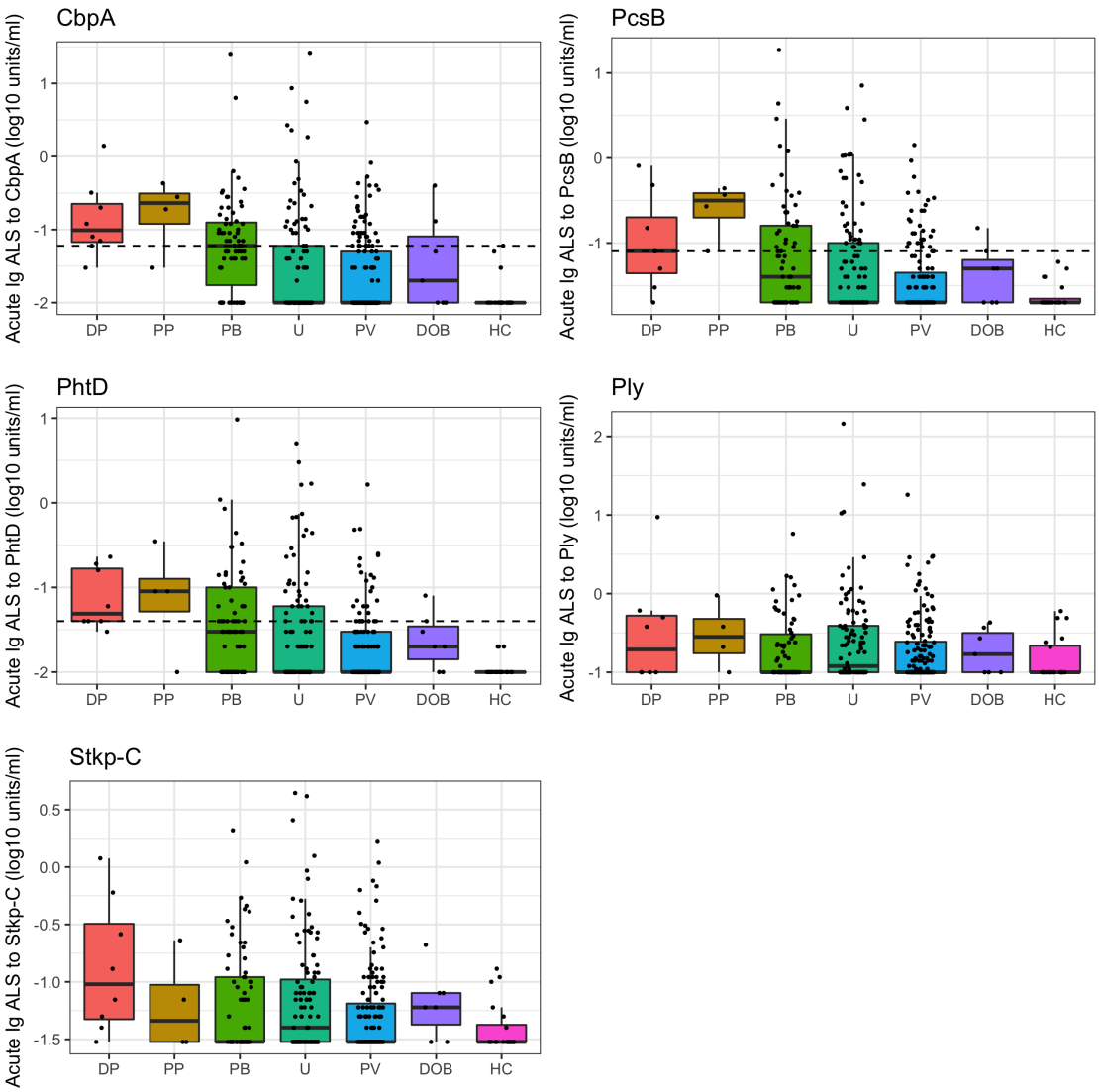


Figure 5.8: ALS to pneumococcal proteins CbpA, PcsB, PhtD, Ply and Stkp-C by pneumonia aetiology comparator standard. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control.

50/267 (19%) in children with CRP concentration <60 mg/l. The association between a positive result of IgG ALS to PhtD and CRP concentration of ≥60 mg/l was significant in univariate logistic regression (p = 0.008), but not when adjusted for age (odds ratio 1.47, 95% CI 0.76, 2.75; p = 0.237). It should be noted that there is significant colinearity between age and CRP concentration in these data, that may be a biological phenomenon, with a higher prevalence of invasive pneumococcal disease and other proxy markers of pneumococcal pneumonia in older children.

Stratification by age $\geq$2 years

The magnitude of ALS concentration to either of CbpA, PcsB and PhtD was correlated with age, as was the prevalence of *definite pneumococcal* pneumonia. In particular, no children <2 years of age on whom ALS to pneumococcal proteins had been measured had *definite pneumococcal* pneumonia (see Table 5.5). In addition, ALS to pneumococcal proteins was consistently in the lower range in children <2 years of age (data not shown). I therefore repeated the analyses (including generation of optimum thresholds) with only samples from children $\geq$2 years of age.

The exclusion of children <2 years of age considerably reduced the ability of IgG ALS to pneumococcal proteins to discriminate between *definite pneumococcal* and *definite other bacterial* pneumonia, although this was considerably limited by just three children with *definite other bacterial* pneumonia remaining in the comparison group. In children $\geq$2 years of age assay of IgG ALS to CbpA had the best diagnostic accuracy, with a sensitivity of 0.875 and specificity of 0.667 for discriminating between *definite pneumococcal* and *definite other bacterial* pneumonia. IgG ALS to other pneumococcal proteins assayed had an area-under-the-curve between 0.5 and 0.667 (i.e. marginally better than chance) for this comparison (Figures 5.9 and 5.10).

5.4 Discussion

This study describes the enrolment and clinical characteristics of 348 children aged 2 months to 14 years of age with a clinical diagnosis of pneumonia, and 20 local healthy control children, at Patan Hospital, Kathmandu Valley. I investigated a recently developed technique, assay of ALS[138], as a diagnostic test pneumococcal infection. These data represent the largest study of the utility of assay of ALS for the diagnosis of infectious disease, and the only study of its utility for the aetiological diagnosis of childhood pneumonia.

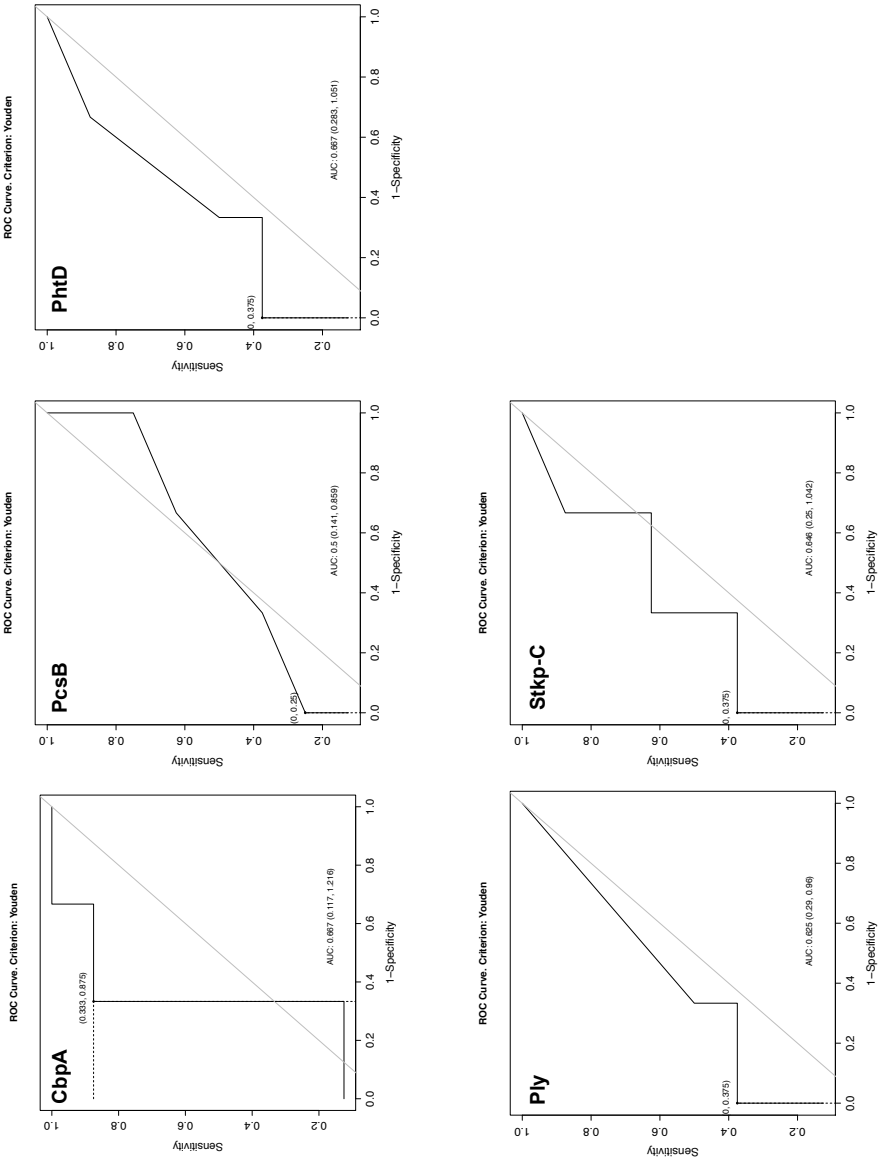


Figure 5.9: ROC curves to determine optimal thresholds for IgG ALS concentrations to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia in children $\geq$2 years of age for five pneumococcal proteins: CbpA, PcsB, PhtD, Ply and Stkp-C.

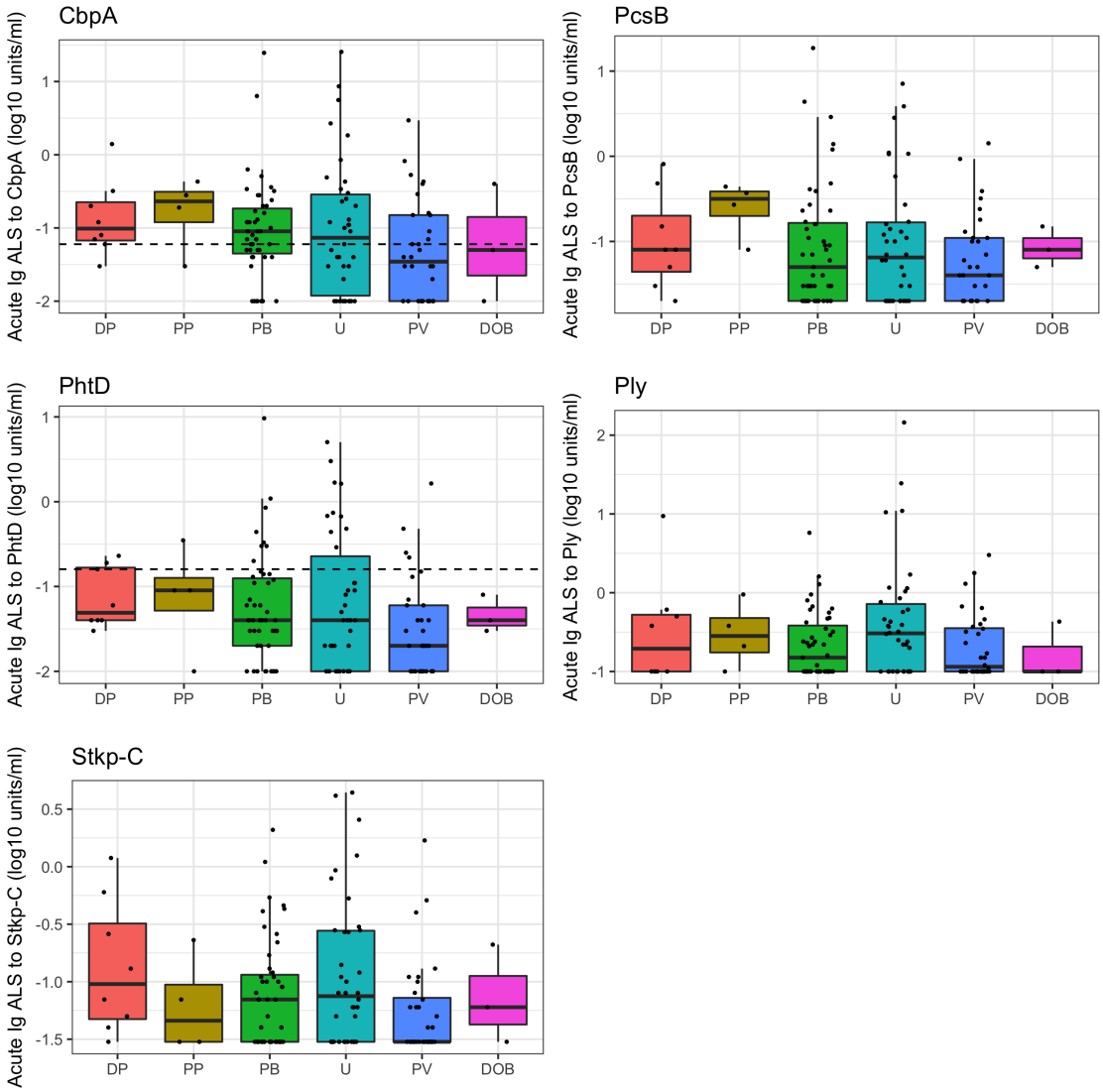


Figure 5.10: ALS to pneumococcal proteins CbpA, PcsB, PhtD, Ply and Stkp-C by pneumonia aetiology comparator standard in children ≥2 years of age. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control.

5.4.1 Pneumonia aetiology comparator standards

The ALS assay is hypothesised to be antigen-specific and time-specific, since it is based on detection and quantitation of antibodies from a population of ASCs that only circulate transiently during the course of an infection (or following vaccination)[121, 262]. To aid interpretation of results, in the absence of a gold standard for the diagnosis of pneumococcal infection, I developed pneumonia aetiology comparator standards *a priori* based on plasma CRP concentrations, NP carriage of serotype 1 pneumococci,

chest radiographs, and blood neutrophil count, as well as enrolment of healthy control children. On the assumption of a low sensitivity of blood culture for pneumococcal pneumonia[2, 3, 33], I hypothesised that there would be a proportion of children with otherwise undiagnosed pneumococcal pneumonia in the comparator standards *probable pneumococcal* pneumonia, *probable bacterial* pneumonia and *unknown* – and possibly *probable viral* pneumonia. I also hypothesised that this proportion would decrease stepwise across these comparator standards (as ordered), and that children with *definite other bacterial* pneumonia and healthy control children did not have pneumococcal infection.

5.4.2 IgG ALS has low specificity and sensitivity for pneumococcal capsular polysaccharides in children with pneumonia

In a set of samples from 100 children with pneumonia there was no association between concentrations of IgG ALS to any capsular polysaccharide and comparator standards for pneumonia aetiology. Indeed, assay of IgG ALS to specific capsular polysaccharides was relatively insensitive and non-specific, even in children with *definite pneumococcal* pneumonia. PS3 was the predominant antibody in ALS measured by FMIA, even in children with invasive pneumococcal disease due to *S. pneumoniae* of a different serotype. IgG ALS concentrations to other capsular polysaccharides were closely correlated. These results were also noted in the WHO-validated Pn PS ELISA in the 2/10 samples that detected IgG over the limit of quantitation for any capsular polysaccharide, and therefore does not appear to be an artefact of the FMIA method.

Although PS3 is known to have a relatively simple molecular structure[42], it is a very rare cause of childhood invasive pneumococcal disease, and is rarely isolated from NP specimens in children with pneumonia, at Patan Hospital (see Chapter 3). A recent study of the IgG response to pneumococcal pneumonia (in serum samples) did not identify a predominance of PS3 antibody[107]. The observation that concentrations of IgG to PS3 were associated with raised concentrations to other pneumococcal capsular polysaccharides, and to Hib polysaccharide suggests that ASCs in the blood of

children with pneumonia may be producing non-specific antibodies to polysaccharide antigens. One might hypothesise that ASCs in ALS from children with acute pneumonia are enriched for recently differentiated plasmablasts that have not undergone affinity maturation. In contrast, serological testing for pneumonia aetiology uses convalescent serum/plasma, where detectable antibodies are derived from mature plasma cells in secondary lymphoid organs and the bone marrow (see Chapter 1).

5.4.3 IgG ALS may discriminate between pneumonia aetiology comparator standards

I found that IgG ALS to the 5 pneumococcal proteins was significantly higher in children with *definite pneumococcal* pneumonia (n = 8) vs healthy control children (n = 20), and that there was a significant trend across all comparator standards for all 5 proteins. In a more rigorous comparison, IgG ALS to PhtD (the "best-performing" antigen) was also significantly higher in children with *definite pneumococcal* than *definite other bacterial* pneumonia (n = 7), and there was a significant reduction in the proportion of children with a positive result of IgG ALS to PhtD across the comparator standards, *probable pneumococcal* pneumonia, *probable bacterial* pneumonia, *unknown* and *probable viral* pneumonia. Concentration of IgG ALS to PhtD (either as a continuous variable, or as a threshold) was also significantly associated with proxy markers of pneumococcal pneumonia: radiographic endpoint consolidation on chest radiograph (including when adjusted for age of child); and plasma CRP concentration (although this latter association was non-significant when adjusted for age). This suggests that IgG ALS to PhtD may be usefully explored as a test to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia.

In this dataset, *definite pneumococcal* pneumonia was only detected in children $\geq$2 years of age. Age may also significantly affect the ability of children to produce an ASC-mediated response to pneumococcal polysaccharide or protein antigens as a result of immature T cell-mediated immunity in this age group. When analyses were limited to children $\geq$2 years of age the ability of IgG ALS to discriminate between *definite*

pneumococcal and *definite other bacterial* pneumonia was reduced. Here IgG ALS to CbpA was the best-performing antigen with a sensitivity of 88% and specificity of 67% for pneumococcal pneumonia – although the number of children with *definite other bacterial* pneumonia was severely limited (n = 3).

5.4.4 Assay of ALS as a novel diagnostic test for pneumonia aetiology?

Assay of ALS was initially developed and trialled for the diagnosis of tuberculosis in adults, in limited case series with a high pre-test probability for a "positive" result. In these case series, using bacille Calmette-Guérin as antigen, sensitivity of IgG ALS generally been $\geq$90%, and specificity $\sim$80–90% at discriminating tuberculosis from other pulmonary infections or healthy controls[147–150, 152] (see Table 1.3.2). The diagnosis of tuberculosis in adults using the host immune response has several advantages over the diagnosis of pneumonia in children. Tuberculosis is a chronic disease, with antigen-specific ASCs in circulation during the course of illness[265], in contrast to acute pneumonia, where the window of detection of antigen-specific ASCs is likely to be much shorter[121, 262]. Detection of an immune response is likely to be generally easier in adults and older children, following the maturation of T cell-mediated immunity, and is already proven in principle with the widespread use of interferon-γ release assays. Finally, the use of case series with a high pre-test probability for the diagnosis of tuberculosis increases the measured sensitivity and specificity of the test. Assay of IgG ALS has also been trialled for the diagnosis of typhoid and paratyphoid fever (enteric fever) in adults and older children in Bangladesh[145, 146]. In these studies adults were only enrolled if symptomatic for 3–7 days, and $\sim$25% were positive by blood culture. Using typhoid membrane preparation as antigen, sensitivity of IgG ALS was $\sim$100% and specificity $\sim$75–97% for the diagnosis of enteric fever vs adults with non-typhoid febrile illness.

In contrast, my results showed a marginal ability to discriminate between pneumococcal pneumonia and pneumonia caused by other bacterial or viral infections using assay of IgG ALS to pneumococcal proteins. If reproducible, this represents an improvement on current state-of-the-art microbiological techniques (see Table 1.2). Here, I assessed an entire

case series of infants and children $\geq$2 months of age with all-cause pneumonia where the pre-test probability for pneumococcal pneumonia is unknown. In addition, children with pneumonia may have had a shorter duration of infection than adults previously assessed for tuberculosis and enteric fever, which may have meant blood samples were taken prior to the circulation of antigen-specific ASCs[121]. An increased understanding of the kinetics of ASC responses in peripheral blood during infection may be helpful to optimise time points for the sampling of ASCs during infection[121].

5.4.5 Limitations and summary

Despite the relatively large size of this case series, the low proportion of children with *definite pneumococcal* (or *definite other bacterial*) pneumonia limits the extent to which these data are robust. It also limits the power of the analyses to detect real differences between children with textitdefinite pneumococcal and *definite other bacterial* pneumonia. For example, additional cases of children with *definite pneumococcal* (and *definite other bacterial*) pneumonia would permit rigorous examination of training and test sets of data. It is unknown if acquisition of NP carriage of a pneumococcal serotype elicits a detectable ASC response in peripheral blood, but if so, this may reduce the specificity of tests of IgG ALS to pneumococcal antigens (under the assumption that the acquisition was not causal to pneumonia). This study was not able to investigate this hypothesis. An additional source of error is the variability of the childhood immune response. In the absence of prior data for ALS assays in children under 2 years of age, I attempted to account for this through a secondary analysis restricted to children greater than 2 years of age.

In summary, I present the design, methods and results for a study of the use of assay of IgG ALS to pneumococcal antigens to identify cases of pneumococcal infection in children aged 2 months to 14 years admitted to hospital in Nepal with a clinical diagnosis of pneumonia. Assay of IgG ALS to pneumococcal proteins, and in particular PhtD, was able to discriminate between *definite pneumococcal* pneumonia and *definite other bacterial* pneumonia with a sensitivity of 88% and specificity of 71% in this small number

of cases. However, when the analyses were limited to children $\geq$2 years of age specificity was reduced.

Future approaches may consider testing ALS against a broader array of pneumococcal protein antigens, and possibly integrating the results to individual antigens (potentially from a broad array of pathogens) using techniques analogous to those developed for the analysis of gene expression microarrays. Assay of ALS to pathogen antigens may also be integrated with a number of diagnostic methods based on proxy markers of pneumonia aetiology, microbiological techniques, and the host immune response (such as assay of ALS, detection of antigen-specific ASCs using ELISpot, serology, and differential gene expression) to predict the aetiology of pneumonia in a given patient. The statistical basis for this has recently been outlined by the Pneumonia Etiology Research for Child Health (PERCH) study team[74].

6

Transcriptomic investigation of childhood pneumonia

6.1 Introduction

Current methods to detect the bacteria or viruses causing pneumonia in children either lack sensitivity (e.g. culture of blood) or specificity (e.g. molecular detection of potential pathogens from the nasopharynx[2]; see Chapter 1). Intensive efforts to integrate results from a wide range of methods for estimating the aetiology of childhood pneumonia in the Pneumonia Etiology Research for Child Health (PERCH) study have highlighted the need for better diagnostic tests. An ideal diagnostic test would be useful both for aetiological diagnosis in individuals (to potentially determine treatment strategy) and for assessment of public health measures (such as in the context of vaccine introduction).

Investigation of the host response to infection may be a useful alternative to clarify the aetiology of childhood pneumonia, and in particular to discriminate between healthy

carriage of potential pathogen and disease. Methods to detect the antibody-secreting cell (ASC) response to childhood infections using either assay of ALS or ELISpot may have utility, but require considerable further investigation (see Chapters 1 and 5). In addition, the ASC response to infection takes several days following onset of disease to be readily detectable in blood, which limits its use to determine treatment strategy[121]. Although the ability to screen antibodies against arrays of pathogen antigens is rapidly increasing[276], studies of antibody-secreting cells are usually limited to identifying antibodies or B cell receptors to pre-supposed antigens. The utility of serological studies for the aetiological diagnosis of childhood pneumonia is uncertain[107, 108], but a convalescent sample is a prerequisite for the diagnosis of almost all pathogens.

An alternative approach is to measure the differential expression of genes across the human transcriptome in the relatively small number of children with pneumonia of known aetiology. This can be used to develop a "host transcriptomic signature" for a given pathogen associated with pneumonia (e.g. *Streptococcus pneumoniae*). If a transcriptomic signature accurately discriminates between patients with known aetiology, it could be modified into a diagnostic test for patients with an unknown pneumonia aetiology[112, 185, 198, 277, 278]. A transcriptomic signature of pneumonia aetiology may have several advantages over diagnostic methods based on direct identification of pathogens (see Chapter 1), or diagnostic methods based on antibodies, or ASCs. RNA is transcribed prior to the translation of proteins, which may increase sensitivity of a test in early disease in comparison to ASC-based methods. The transient nature of RNA may increase temporal specificity in comparison to serological testing (where pre-existing pathogen-specific IgG renders a single acute serological sample non-specific). Finally, transcriptomic methods are relatively unbiased and may therefore describe novel characteristics of the immune response to pneumonia.

As detailed in Chapter 1, previous work has shown gene expression data from microarray experiments to discriminate between aetiologies of infectious disease syndromes in children, including differentiating bacterial from viral infections[112, 185, 277, 278], between respiratory viral infections[198], between classes of bacterial pathogens[189] and tuberculosis from other bacterial and viral infections[190, 191]. Direct sequencing of

RNA is increasingly used for the generation of transcriptomic data. The resulting data have a greater dynamic range than that of microarray experiments (which are limited by fluorescence values for individual probes), and may be explored for differential expression of different gene isoforms[177].

Here, I investigated if differential gene expression, derived from RNA sequencing of blood samples taken during acute disease, could discriminate between aetiologies of pneumonia in children in Nepal. I aimed to generate an accurate transcriptomic signature of pneumococcal pneumonia. A secondary aim was to develop an accurate test to discriminate between bacterial and viral pneumonia. I also hypothesised that, given the limitations of studying pneumococcal vaccine impact described previously (Chapter 4), a transcriptomic signature of pneumococcal pneumonia, in a high throughput manner with a different technology platform, could be used to assess the impact of 10-valent PCV on childhood pneumococcal pneumonia in Nepal.

6.2 Methods

Ethics statement

Parents of all children enrolled to the study gave witnessed, written, informed consent before enrolment of their child into the study. The Nepal Health Research Council (reference 286/2014) and the Oxford Tropical Research Ethics Committee (reference 09/15) approved the study protocols.

Study site and population

Details of the study site are described in Section 2.1.

Summary of clinical, radiological, laboratory and microbiological methods

The background to the studies of childhood pneumonia at Patan Hospital is described in Section 2.3.

6.2.1 Enrolment and sampling

The investigation of the host transcriptome for the aetiological diagnosis of pneumonia, and investigation of pathophysiology, commenced enrolment on 26th February 2015. This investigation formed part of the PANDIT Study. Here I present data on patients enrolled until the end of December 2016.

Children $\geq$2 months (60 days) of age to 14 years of age who were being admitted to Patan Hospital with a clinical diagnosis of pneumonia were eligible for this study. From September 2015 children $\geq$2 months (60 days) of age and up to 14 years of age who were diagnosed with clinical pneumonia at Patan Hospital, but not admitted to hospital, were also deemed eligible for enrolment. The clinical diagnosis of pneumonia was made by the admitting (or reviewing) paediatric team and the decision to admit to hospital was based on clinical discretion and the WHO *IMCI* guidelines[94]. All children enrolled in the study had a chest radiograph and a single NP swab taken for culture of pneumococci. Enrolled children also had blood taken aseptically for blood culture, full blood count, white cell differential and measurement of plasma CRP concentration as described in Section 2.3.

In addition, 50 children from the community ("healthy controls") were enrolled into the PANDIT study to provide an additional comparator standard for the analysis of the host transcriptome in childhood pneumonia. These were all local children, 10 months of age, who had received three doses of 10-valent PCV, with at least four weeks between the third dose of PCV (or any other vaccine) and blood sampling. The enrolment strategy for the study is detailed in Section 5.2.

Samples for host transcriptomic investigation

On enrolment to the PANDIT Study, an additional 3 ml of blood was taken, either at the same venepuncture as sampling for blood culture, full blood count and white cell differential, or within 48 hours of admission. 1 ml of blood for transcriptomic analysis was immediately inoculated into RNA stabilising reagents in a purpose-designed collection tube (Tempus Blood RNA Tube, Applied Biosystems, USA) and stored at -80°C within 4 hours of sampling. The remaining 2 ml of blood was used for the generation of ALS

(see Chapter 5). All children enrolled in the study were offered the opportunity to return to Patan Hospital to donate a further sample of 3 ml of blood at 6–8 weeks following their initial admission. Healthy controls had blood taken at a specific research venepuncture but this was processed as detailed above.

6.2.2 Development of comparator standards for pneumonia aetiology

In the absence of a "gold standard" against which to compare the results of transcriptomic analyses for the aetiology of childhood pneumonia, I developed a series of comparator standards. These pneumonia aetiology comparator standards were all developed *a priori*. The comparator standards were hypothesised to provide a trend from a high proportion of children with pneumococcal pneumonia to a low proportion of childhood pneumonia.

The comparator standards were, "*definite pneumococcal*" pneumonia, "*probable pneumococcal*" pneumonia, "*probable bacterial*" pneumonia, "*unknown*" pneumonia, "*probable viral*" pneumonia, "*definite other bacterial*" pneumonia and healthy control. For clarity, these comparator standards are italicised throughout this chapter. The rationale behind these comparator standards is described in Chapter 5, using supporting data described in Chapter 3.

6.2.3 Peripheral blood RNA extraction and quantification

The methods of blood sampling, RNA stabilisation, RNA extraction, library preparation and sequencing are detailed in Section 2.5.1. In brief, samples for host transcriptomics were shipped to the UK on dry ice and RNA was extracted from Tempus Blood RNA Tubes using the Tempus Spin Isolation Kit, according to the manufacturer's protocol[219]. Library preparation, depletion of ribosomal and globin RNA, and sequencing of RNA from selected samples was undertaken at the WTCHG, University of Oxford, using Illumina HiSeq4000 sequencing to generate 75 bp paired end reads.

All available samples were selected for RNA sequencing from the comparator standards *definite pneumococcal*, *probable pneumococcal* and *definite other bacterial* pneumonia. Samples from comparator standards *probable bacterial*, *unknown*, *probable viral* pneumonia, and healthy controls, were chosen by computer-generated random selection of all pairs of samples (i.e. from children with an acute and convalescent sample), followed by computer-generated random selection of single acute samples until the pre-defined number of samples were selected for each comparator standard. The predefined number of samples for the comparator standards *probable bacterial*, *unknown*, *probable viral* pneumonia, and healthy controls were derived from estimates of the likely diversity of pathogenic organisms, and variance of differential gene expression within each group.

6.2.4 Bioinformatic approaches

Bioinformatic methods are detailed in Section 2.5.2. In brief, 75 bp paired end reads were aligned to the human genome (build GRCh38) using the alignment tool Bowtie2 as part of the spliced junction mapping package, Tophat[178]. Following alignment and filtering to remove duplicated reads, I used HTSeq[179] to generate counts of reads mapped to each gene isoform. These data were imported into the R statistical environment for further analyses. Only protein coding genes from autosomes were used for these analyses, with sequences mapping to other parts of the human genome removed. Gene transcription levels were normalised using trimmed means of M-values[222], and genes that were only expressed in less than 8 samples (the smallest biologically relevant group – *definite pneumococcal* pneumonia) were removed from analyses.

Differential gene expression

Differential gene expression was calculated using the R package $edgeR$ (see Section 2.5.2) for the following comparisons:

1. Grouped acute samples from children with pneumonia (any comparator standard) vs grouped healthy controls (Figure 6.2, comparison 1).

2. Paired acute samples from children with acute pneumonia* vs convalescence (Figure 6.2, comparison 2).

*For this comparison, "acute pneumonia" represents the comparator standards *definite pneumococcal*, *probable pneumococcal*, *probable bacterial*, *probable viral* and *definite other bacterial* pneumonia in comparison to samples taken at convalescence from these children. A subsidiary analysis was undertaken on paired acute and convalescent samples from children with *definite pneumococcal* and *probable pneumococcal* pneumonia only.

3. Grouped acute samples from children with *definite pneumococcal*, *probable pneumococcal* or *definite other bacterial* pneumonia (together forming "bacterial" pneumonia) vs *probable viral* pneumonia (termed "viral" pneumonia when defined in this comparison; Figure 6.2, comparison 3).

4. Grouped acute samples from children with *definite pneumococcal* and *probable pneumococcal* pneumonia vs *definite other bacterial* pneumonia (Figure 6.2, comparison 4).**

**Following inspection of the data, children with *probable pneumococcal* pneumonia were grouped with children with *definite pneumococcal* pneumonia to form "pneumococcal" pneumonia for comparison 4. Figure 6.2 provides a visual schematic for the comparisons of differentially expressed genes between different groups of children.

I hypothesised that these comparisons between comparator standards could derive a transcriptomic signature of pneumonia (likely to be a highly diverse group, due to differing pathogens and severity of illness) vs a state of health (comparison 1); and a transcriptomic signature of acute pneumonia vs convalescence (comparison 2). I explored comparisons 1 and 2 through the use of volcano plots, heat maps and gene pathway analysis in a descriptive manner.

Discrimination between bacterial and viral pneumonia is of clinical (and public health) significance. To investigate this, I used descriptive analyses of differential gene expression between children with bacterial pneumonia (as defined above) vs viral pneumonia (comparison 3). I then investigated the ability of this bacterial transcriptomic signature to be developed into a "Bacterial Gene Score" to to discriminate between bacterial and viral pneumonia. This Bacterial Gene Score was created by the summation of the normalised differential gene expression value for each child. For example, if gene A was upregulated by a value of 10, gene B was upregulated by a value of 20, and gene C was downregulated by a value of 5, the gene score would equal 10 + 20 - 5 = 25. The ability of the Bacterial Gene Score to discriminate between *bacterial* and viral pneumonia was then assessed using receiver operating characteristic (ROC) curves (according to the Youden index, using $OptimalCutpoint$[272], see Section 5.2) and an optimal threshold for the Bacterial Gene Score was determined. This threshold was then applied retrospectively to the Bacterial Gene Score corresponding to RNA sequencing data from all children to estimate the proportion of children with bacterial pneumonia in each comparator standard.

I used descriptive analyses of differential gene expression between children with pneumococcal pneumonia and *definite other bacterial* pneumonia (comparison 4). I then investigated the ability of this pneumococcal transcriptomic signature to be developed into a "Pneumococcal Gene Score" to discriminate between pneumococcal and *definite other bacterial* pneumonia, analogous to the gene score developed for comparison 3. Similarly, I used ROC curve analysis to to determine an optimal threshold for the Pneumococcal Gene Score. This threshold was then applied retrospectively to the Pneumococcal Gene Score corresponding to RNA sequencing data from all children to estimate the proportion of children with pneumococcal pneumonia in each comparator standard.

6.3 Results

6.3.1 Enrolment, sampling and comparator standards

During the study period, 610 children aged 2 months to 14 years were admitted to Patan Hospital with a clinical diagnosis of pneumonia; 514 (84%) were enrolled into the study. An additional 40 children with a clinical diagnosis of pneumonia made at Patan Hospital, but not admitted to hospital, were also enrolled, totalling 554 children. No children with pneumonia died during the study period. In addition, 50 healthy control children were enrolled. These control children were all 10 months of age, without intercurrent illness and had all received 3 doses of 10-valent PCV with the most recent dose of any vaccine >28 days prior to sampling of blood.

Sample availability from these children is described in Figure 6.1. 23 acute samples that were initially available were rejected due to an RNA concentration <750 ng/μl following extraction of RNA and did not go forwards to sample selection. The methods of sample selection for RNA sequencing is detailed in the Methods section.

RNA sequencing was performed on acute samples from seven children with *definite pneumococcal*, four children with *probable pneumococcal*, 40 children with *probable bacterial*, 26 children with *unknown*, 25 children with *probable viral* and eight children with *definite other bacterial* pneumonia, and 25 healthy controls. Transcriptomic data from paired convalescent samples was available from four children with *definite pneumococcal*, two children with *probable pneumococcal*, 15 children with *probable bacterial*, 16 children with *unknown*, 16 children with *probable viral* and four children with *definite other bacterial* pneumonia (Figure 6.1).

For comparisons of bacterial pneumonia vs viral pneumonia (comparison 3, Figure 6.2), data from children with *definite pneumococcal* and *probable pneumococcal* and *definite other bacterial* pneumonia are aggregated, totalling 19 children with bacterial pneumonia. For comparisons of pneumococcal pneumonia vs *definite other bacterial* pneumonia, data from children with *definite pneumococcal* and *probable pneumococcal* pneumonia were aggregated, totalling 11 children with pneumococcal pneumonia. As a

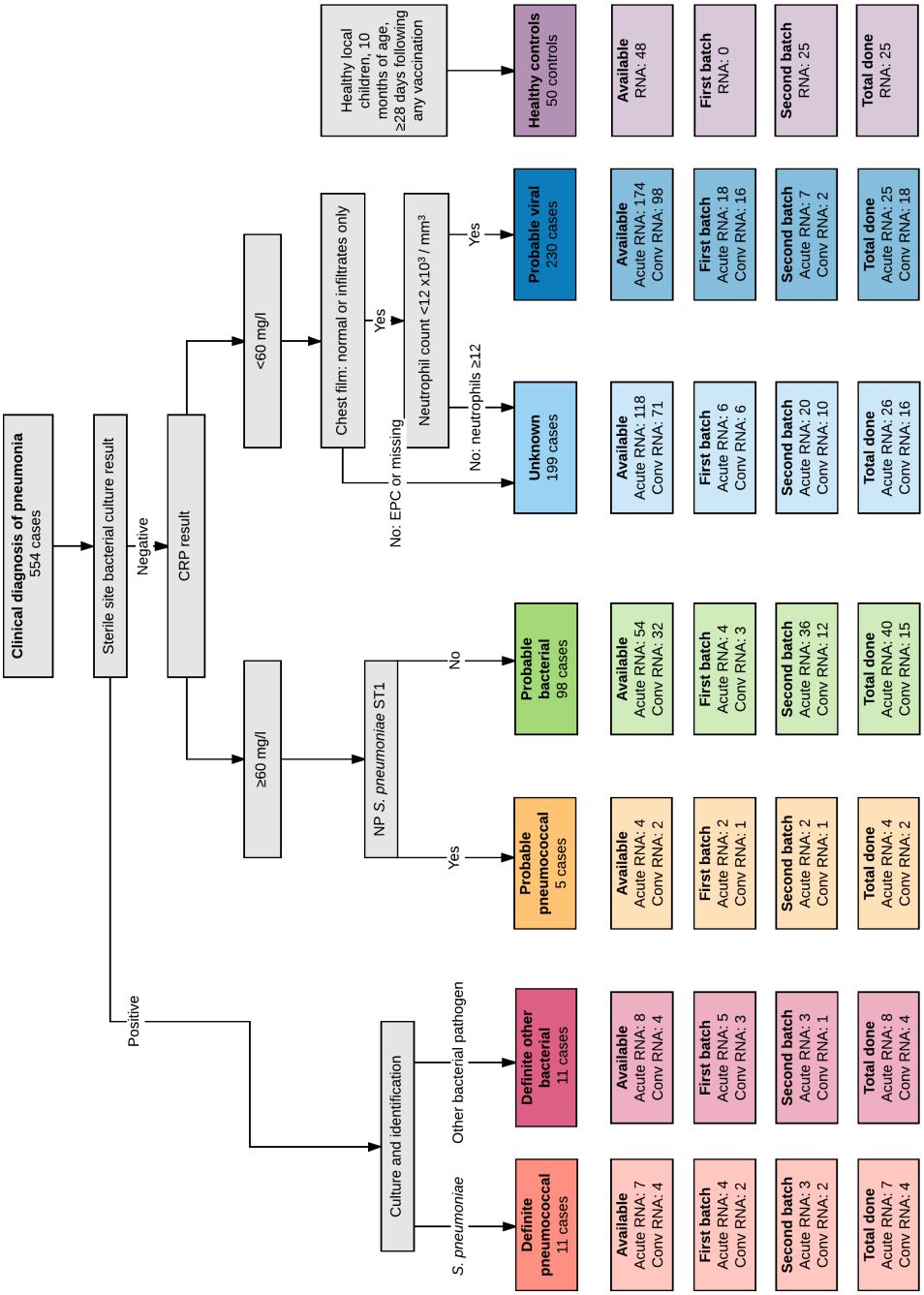


Figure 6.1: Enrolment and sample availability for the transcriptomic investigation of childhood pneumonia.

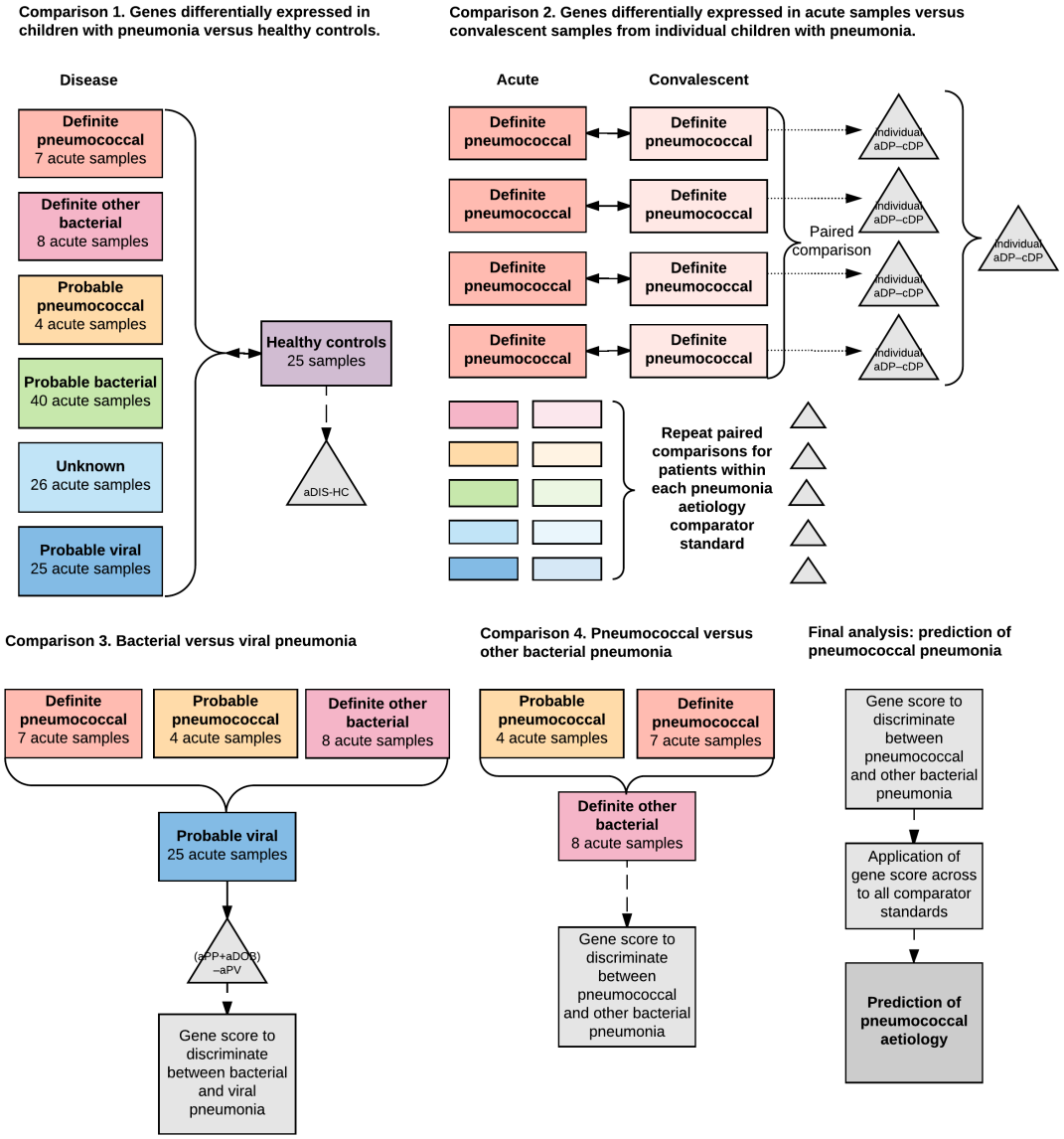


Figure 6.2: Comparisons of differential gene expression to develop transcriptomic signatures for childhood pneumonia across comparator standards of pneumonia aetiology and healthy controls.

group, children with pneumococcal pneumonia displayed similar clinical characteristics, as described here and in previous chapters, independently of their transcriptomic data regardless of whether they were initially classified as *definite pneumococcal* or *probable pneumococcal* pneumonia.

Details on the characteristics of children on whom transcriptomic investigations were undertaken are described in Table 6.1. Children with bacterial pneumonia were more likely to be male than children with viral pneumonia (91% vs 44%, p = 0.024). There was

no significant difference in the sex ratio between children with pneumococcal pneumonia and *definite other bacterial* pneumonia. Children with bacterial pneumonia were older than children with viral pneumonia (median 5.8 years vs 1.1 years, p <0.001), and in turn, children with pneumococcal pneumonia were older than children with *definite other bacterial* pneumonia (median 7.8 years vs 0.8 years, p <0.001). Across all children, the median length of illness was 4.0 days. Children with bacterial pneumonia were unwell for slightly longer than children with viral pneumonia (median 4.0, IQR 3–7, vs 4.0, IQR 2.5–6, p<0.001). Children with pneumococcal pneumonia were unwell for slightly shorter than children with *definite other bacterial* pneumonia (median 4.0, IQR 2.5–5.5, vs 4.0, IQR 3.5–6.25, p = 0.001).

6.3.2 Principal component analyses

Analysis of batch effects and sampling time points

Principal component analyses were examined visually for artefact arising from using two batches of processing for RNA sequencing, before and after normalisation of differential gene expression. There was no discernible batch effect, and batch was not included in models of differential gene expression. Figure 6.3 (upper) shows the principal and secondary components for all samples within the study, as points coloured by batch. Inspection of the principal and secondary components for all samples, coloured by time point, showed overlapping but relatively distinct clusters of samples from children with acute pneumonia (red), convalescence (green) and healthy controls (blue), as show in Figure 6.3 (lower). Convalescent samples clustered closely to healthy controls. Principal component 1 accounted for 28% of the variability in these samples – thus time point of sampling (acute and convalescent and healthy controls) is associated with 28% of the observed variance in differential gene expression in these children.

Comparator standards

Differential gene expression across all acute samples and healthy controls showed overlapping but relatively distinct clusters of samples when coloured by pneumonia aetiology comparator standard (Figure 6.4). Each ellipse represents the 95% confidence

Table 6.1: Characteristics of children on whom transcriptomic investigations (including analysis of differential gene expression) were undertaken, by pneumonia aetiology comparator standard. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control.

	Comparator standard									
	DP	**PP**	**PB**	**U**	**PV**	**DOB**	**HC**	**Pneumonia**	**p (PV vs [a] DP + PP)**	**p (DP + PP vs DOB)**
n (acute samples)	7	4	40	26	25	8	25	110		
n (paired samples)	4	2	15	16	16	4		57		
Male sex	6 (86%)	4 (100%)	23 (58%)	17 (65%)	11 (44%)	4 (50%)		65 (59%)	0.024	0.141
Age in years (median, IQR)	5.8 (4.7, 8.5)	8.4 (7.9, 9.3)	2.2 (1.1, 3.9)	2.0 (0.9, 4.0)	0.8 (0.6, 1.5)	0.8 (0.6, 5.8)		1.8 (0.8, 4.3)	<0.001	<0.001
2 months to <2 years	0	0	19 (48%)	13 (50%)	22 (88%)	5 (63%)		59 (54%)		
2 years to <5 years	2 (29%)	0	12 (30%)	9 (35%)	2 (8%)	0		25 (23%)		
5 years to 14 years	5 (71%)	4 (100%)	9 (23%)	4 (15%)	1 (4%)	3 (38%)		26 (24%)		
Length of illness (median, IQR)	4.0 (2.5, 5.5)	3.5 (2.8, 4.5)	4.0 (3.0, 7.3)	3.0 (2.0, 5.8)	4.0 (3.0, 7.0)	5.0 (3.5, 6.3)		4.0 (3.0, 6.8)	<0.001	<0.001
CRP mg/l (median, IQR)	184 (80, 262)	182 (85, 176)	111 (85, 176)	22 (5.8, 40)	14 (2.4, 18)	15 (11, 112)		49 (14, 114)		
CRP mg/l $\geq$60 mg/l	5 (71%)	4 (100%)	40 (100%)	25 (96%)	0	3 (38%)		52 (47%)		
Neutrophil count x $10^3/mm^3$ (median, IQR)	11 (6.4, 17)	22 (16, 28)	12 (9.4, 16)	9.7 (5.6, 14)	6.3 (4.0, 8.4)	9.7 (6.8, 11)		9.8 (5.8, 14)		
Endpoint consolidation	6 (100%)	3 (75%)	23 (62%)	17 (74%)	0	2 (29%)		51 (46%)		
Pneumococcal carriage	3 (43%)	4 (100%)	15 (38%)	8 (31%)	18 (72%)	2 (25%)		50 (45%)		
Pneumococcal carriage (no antibiotics)	1/3 (33%)	3/3 (100%)	12/24 (50%)	4/13 (31%)	12/14 (86%)	0/2		32/59 (54%)		

[a] p values calculated using χ^2 test without continuity correction for proportions, or the Mann-Whitney U test for ordinal/continuous variables.

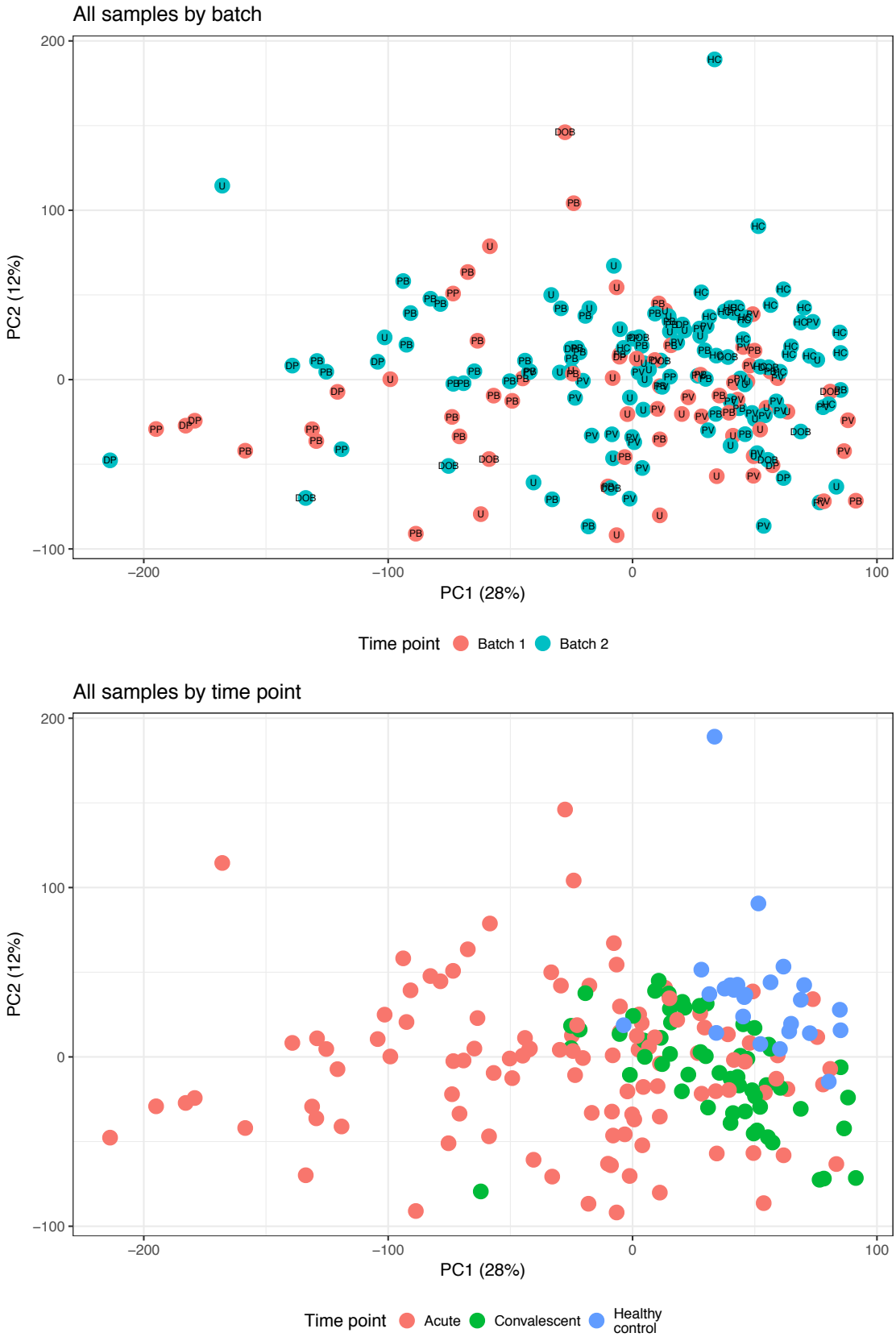


Figure 6.3: Variance in gene expression across cases (acute and convalescent samples) and healthy controls shown by principal component analysis. The upper figure colours samples by sequencing batch and points are annotated with their corresponding pneumonia aetiology comparator standard: DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; HC healthy control. The lower figure colours samples by sampling time point/healthy controls.

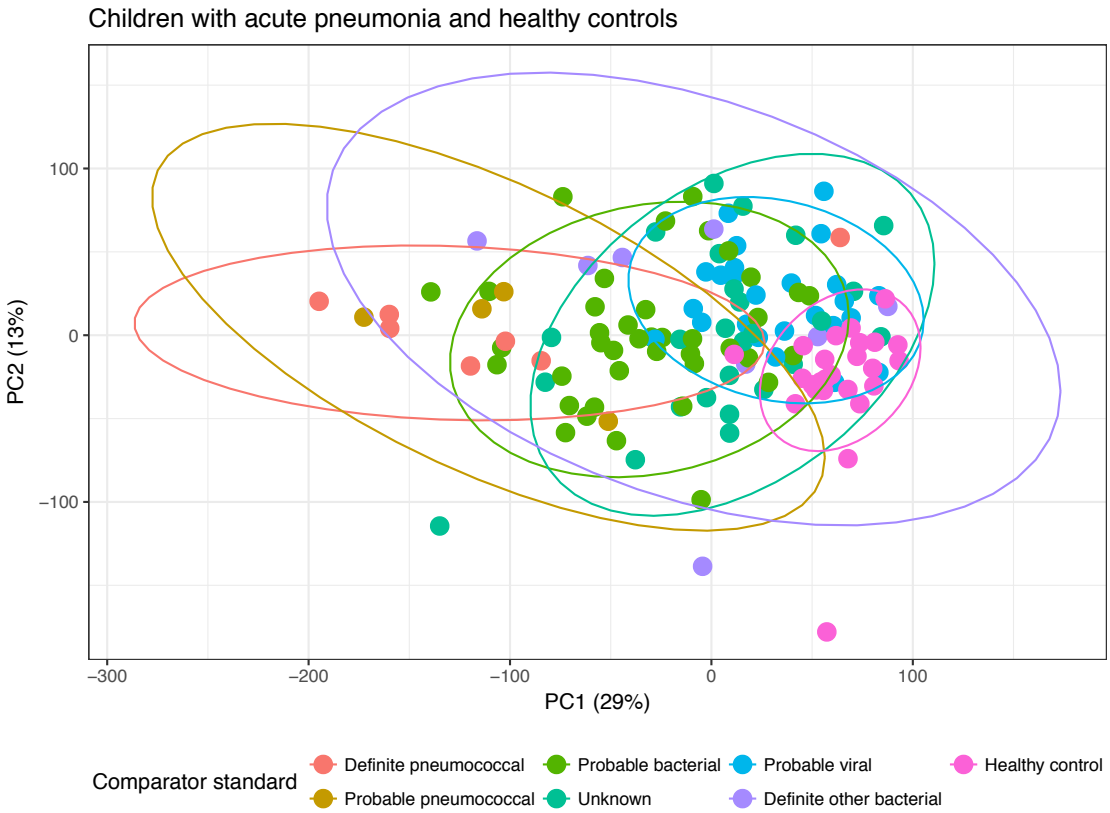


Figure 6.4: Principal component 1 and 2 of all differential gene expression between children with acute pneumonia (all comparator standards) and healthy controls. Each ellipse represents the 95% confidence interval for membership of a group.

interval around the centre of the cluster for each comparator standard. Samples from children with *definite pneumococcal* pneumonia (red) and *probable pneumococcal* pneumonia (gold) clustered relatively closely, with the exception of one outlying sample from a child with pneumococcal pneumonia. Samples from children with *probable bacterial* pneumonia (green) tended to cluster with samples from children with *unknown* pneumonia aetiology (turquoise). Samples from children with *probable viral* pneumonia (blue) clustered intermediate to samples from children with *unknown* aetiology and healthy controls. Samples from children with *definite other bacterial* pneumonia were widely distributed across principal component 1 and 2. To assess the ability of transcriptomic data to discriminate between comparator standards, formal comparisons of comparator standards (or grouped comparator standards) were made. The results from these analyses are described below.

6.3.3 Comparison 1: all-cause pneumonia vs healthy controls

I investigated transcriptomic differences between children with pneumonia vs healthy controls using $edgeR$ to detect differential gene expression as described in Section 2.5.2. Using unsupervised hierarchical clustering and a stringent threshold false discovery rate (FDR) of 0.001, I identified 1290 differentially expressed genes – 1037 genes upregulated in pneumonia, and 253 genes downregulated in pneumonia. Figure 6.5 describes this differential gene expression using a plot of $\log_2$ fold change of gene expression on the x axis, and $\log_{10}$ adjusted p-value (i.e FDR) on the y axis (a "volcano plot"), with each point representing the value for a single gene. Unsupervised clustering of samples from children with pneumonia and healthy controls is shown in a heat map (Figure 6.6). The dendrogram and colour bar illustrate the clustering of samples, with all samples from healthy controls contained within a single central cluster, alongside a proportion of samples from children with pneumonia.

Application of the XGR algorithms to the REACTOME ontology (at FDR = 0.01) revealed the enrichment of a number of upregulated pathways (Figure 6.7). These pathways were almost all inter-related and corresponded to immunological activation. For example, the top 5 most enriched pathways were, neutrophil degranulation, the innate immune response, toll-like receptor signalling and immunoregulatory interactions. At this significance threshold, there was no enrichment of downregulated pathways.

6.3.4 Comparison 2: acute pneumonia vs convalescence

Paired comparisons of the transcriptome between acute pneumonia and convalescence in samples from the same child may reveal the underlying transcriptomic features of pneumonia (and more specifically, pneumococcal pneumonia) with high statistical power. Here, I describe differential gene expression between samples from children with pneumonia and a paired convalescent sample.

Figure 6.8 shows principal components 1 and 2 of the gene expression data for paired samples from children with pneumonia and convalescence. Paired samples are joined

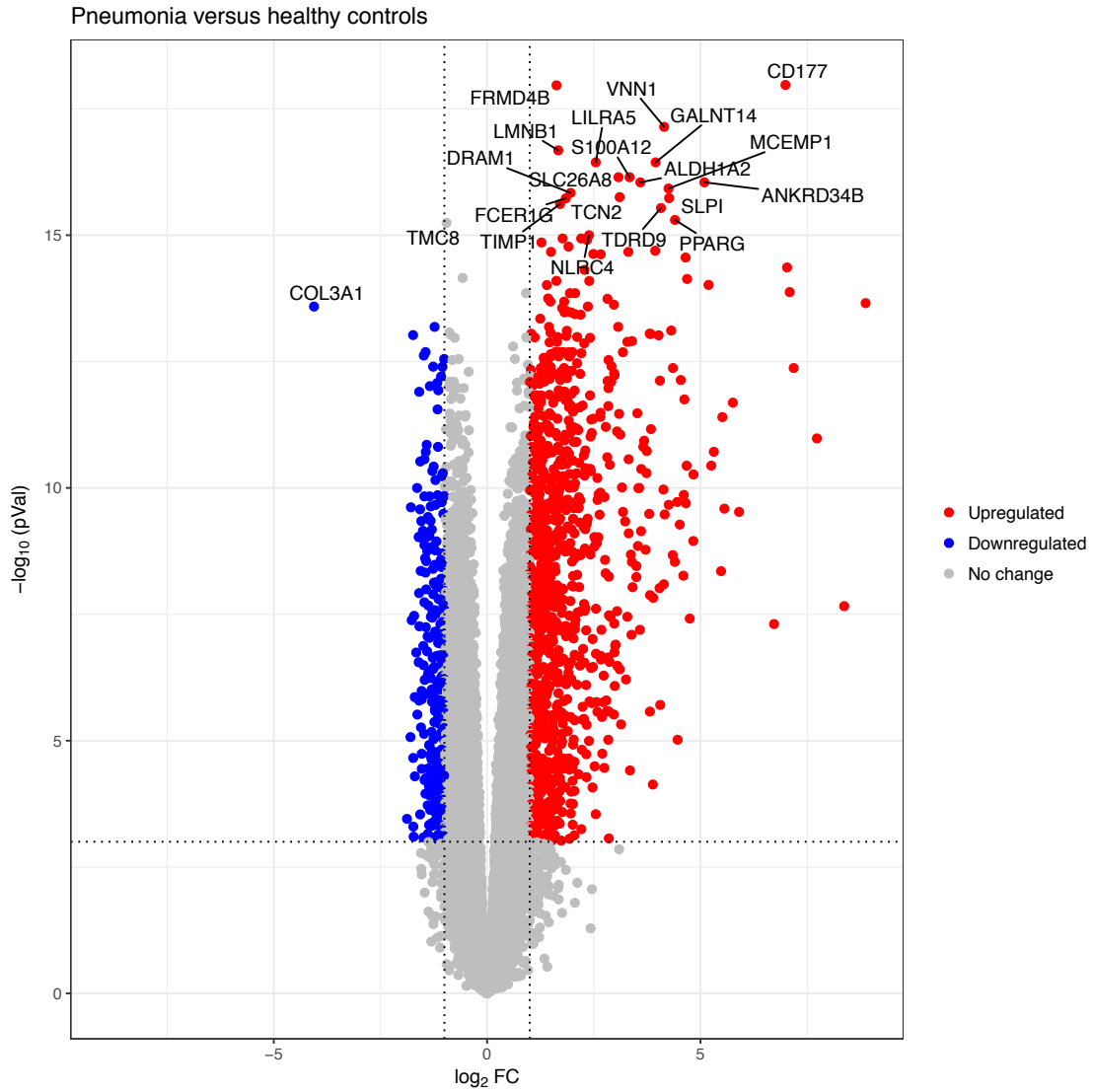


Figure 6.5: Volcano plot of differential gene expression between children with acute pneumonia (any comparator standard) and healthy controls. Each point represents a single gene. Points coloured red indicate upregulated genes (FDR <0.001 and $\log_2$ fold change >2) in children with acute pneumonia vs healthy controls. Points coloured green indicate downregulated genes in children with acute pneumonia. The top most differentially regulated genes are annotated with corresponding gene names.

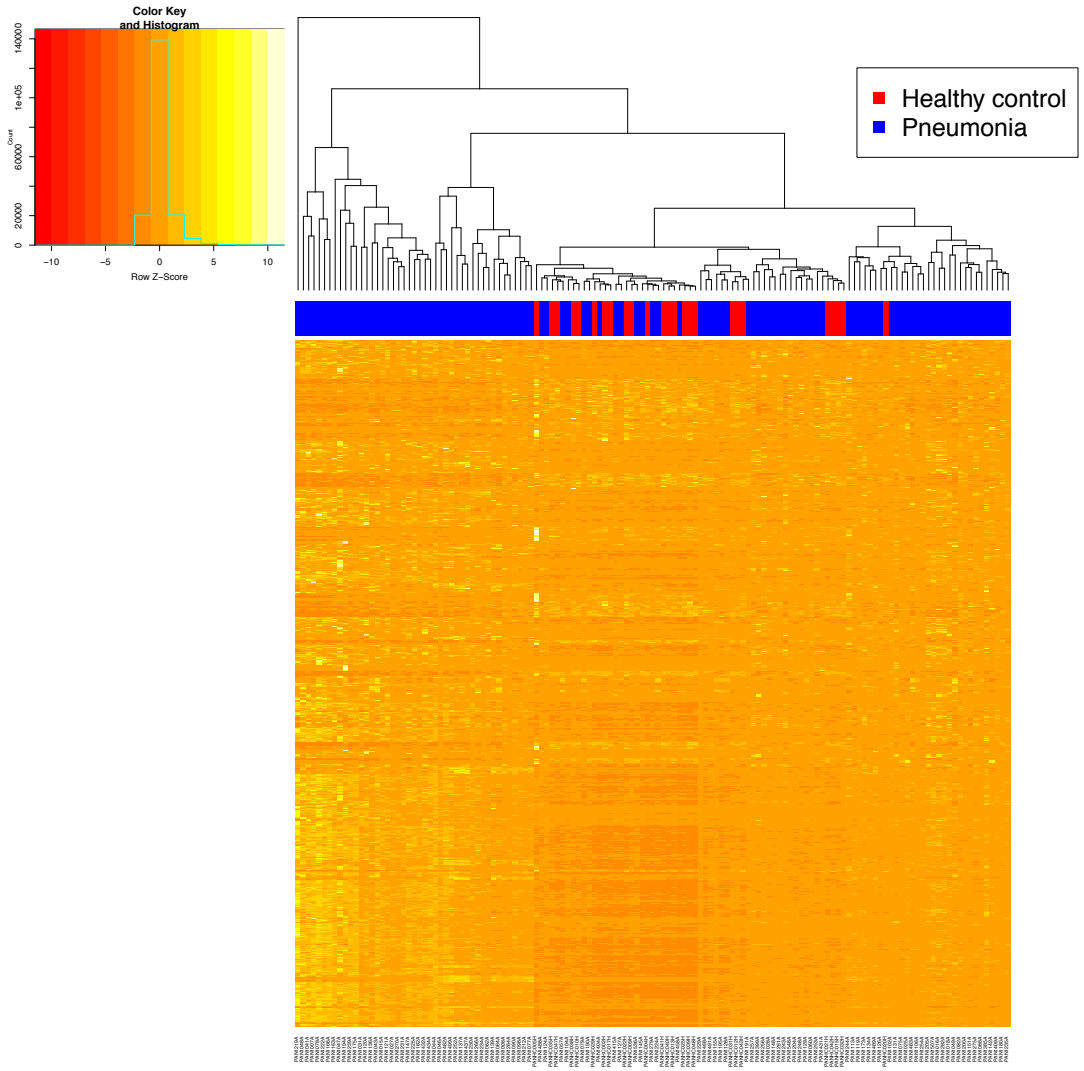


Figure 6.6: Heatmaps of all differentially expressed genes in children with acute pneumonia (of any comparator standard) and healthy control children. Clustering of patients was unsupervised, and provided three main groups (explored in later analyses) with healthy control children all within the central cluster.

by a grey arrow, with the direction of the arrow representing the change between acute pneumonia to convalescence. Acute samples from children with pneumococcal pneumonia and *definite other bacterial* pneumonia tend to be dispersed across the principal components. A number of samples from children with *probable bacterial* pneumonia and *unknown* aetiology are also widely dispersed. Samples from children with *probable viral* pneumonia are less widely dispersed, and plot closely to samples from children in convalescence. This suggests that, at this level of analysis, the transcriptome

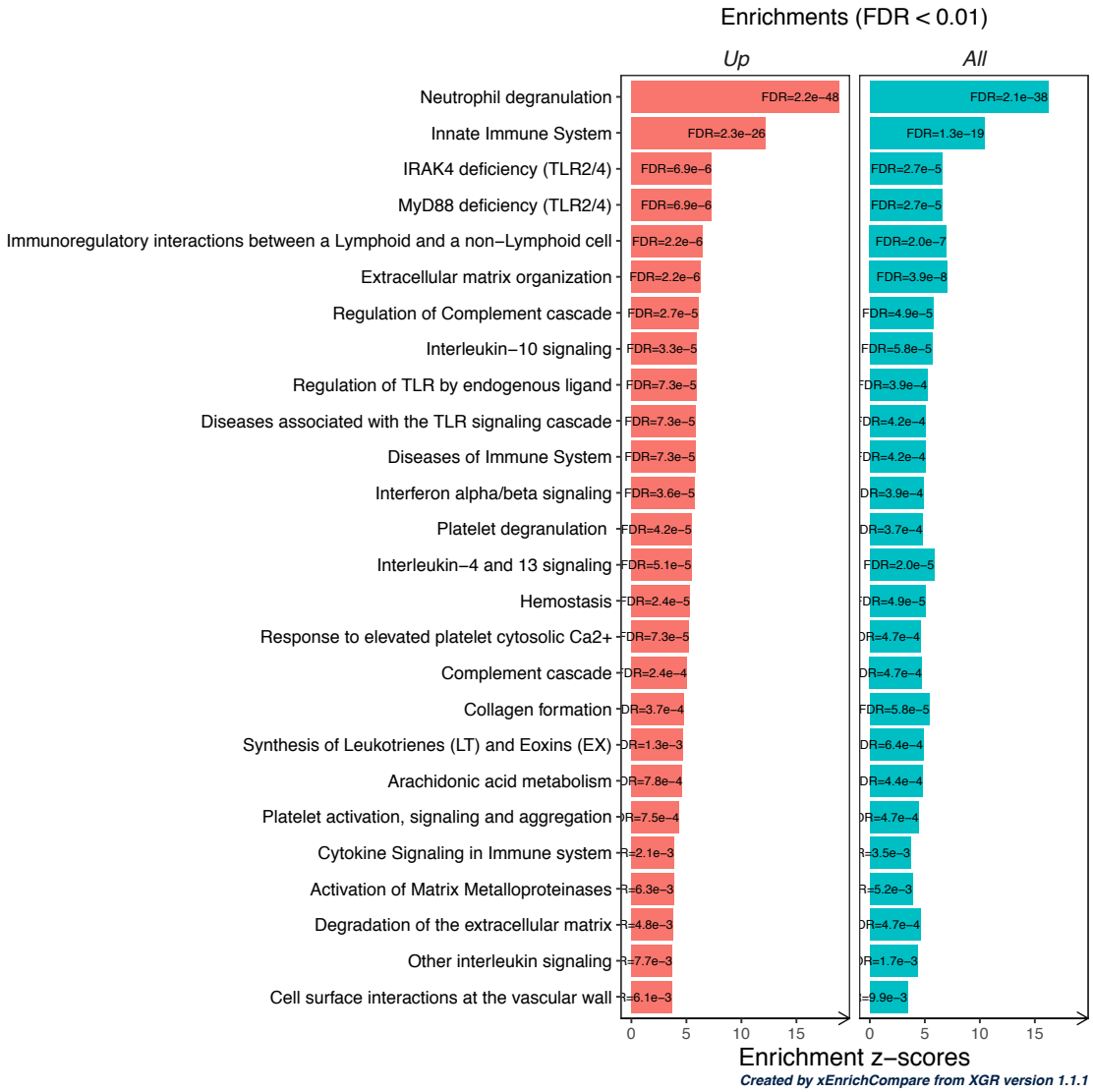


Figure 6.7: Enrichment of gene networks between children with acute pneumonia (any comparator standard) and healthy control children. Networks were derived from the REACTOME ontology using the XGR package in R as detailed in the Methods.

of children in convalescence broadly "returns" to a healthy state 6–8 weeks following disease.

Analysis of differential gene expression between samples from children with acute pneumonia and convalescence showed 637 genes to be differentially expressed: 584 genes were upregulated and 53 genes were downregulated (FDR = 0.001). The differential gene expression between acute pneumonia and convalescence is summarised in Figure 6.9, with the top most differentially regulated genes annotated. Clustering of samples from children with acute pneumonia and convalescence is shown in Figure

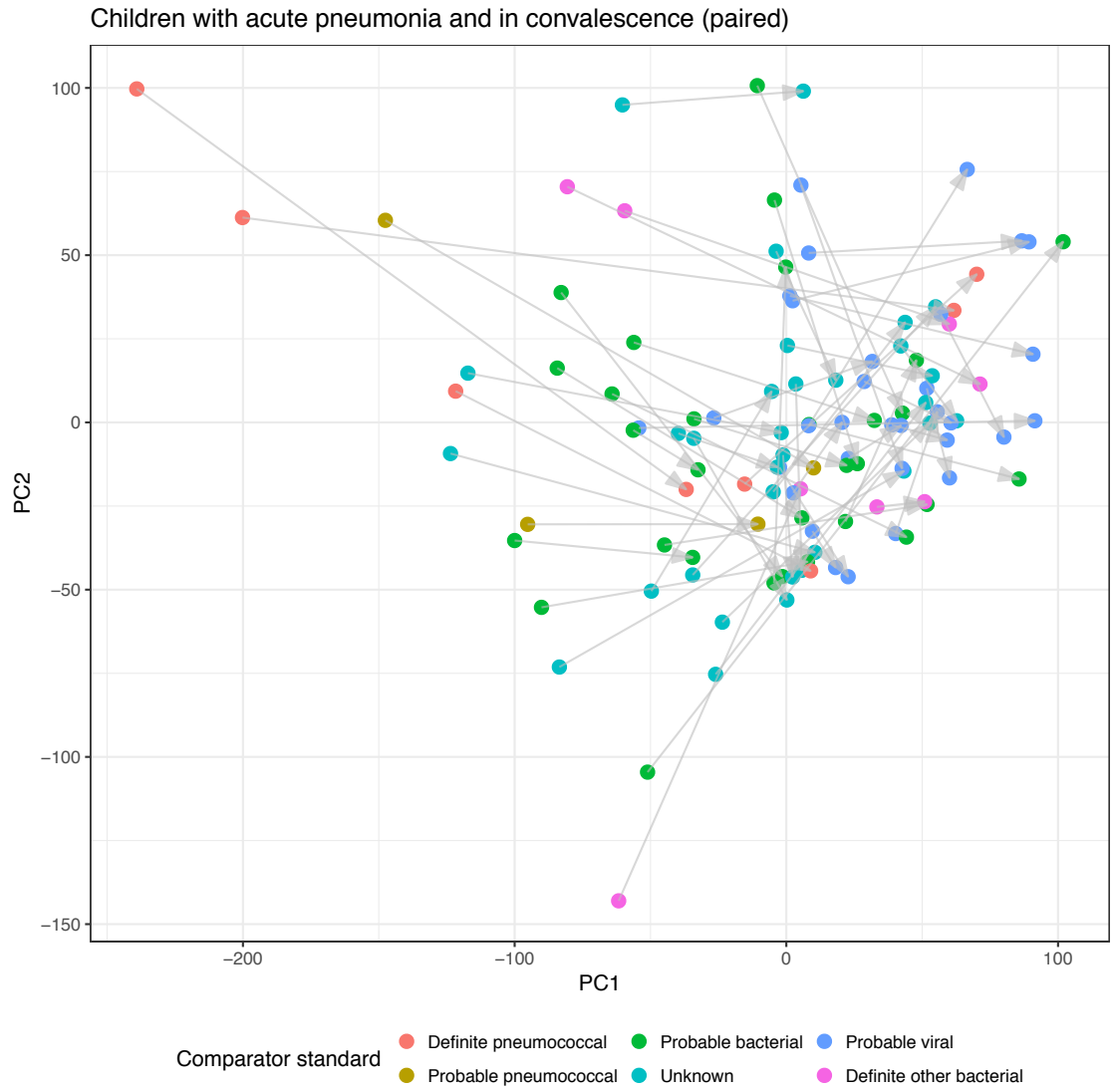


Figure 6.8: Principal component 1 and 2 of all differential gene expression between a child with acute pneumonia, and the same child in convalescence (6–8 weeks following the first sample). Grey arrows link acute and convalescent samples from the same child (with the arrowhead pointing towards convalescent samples).

6.10. There is a trend towards clustering of samples from children with acute pneumonia to the left of the dendrogram, but no clear demarcation between acute pneumonia and convalescence in these samples from children with pneumonia aetiology of any comparator standard. The lack of specific clustering reflects the overlap between clusters of samples visualised in the principal component analysis (Figure 6.8).

Pathways associated with immunological activation were enriched in children with acute pneumonia vs convalescence (Figure 6.11). As with the comparison of children with

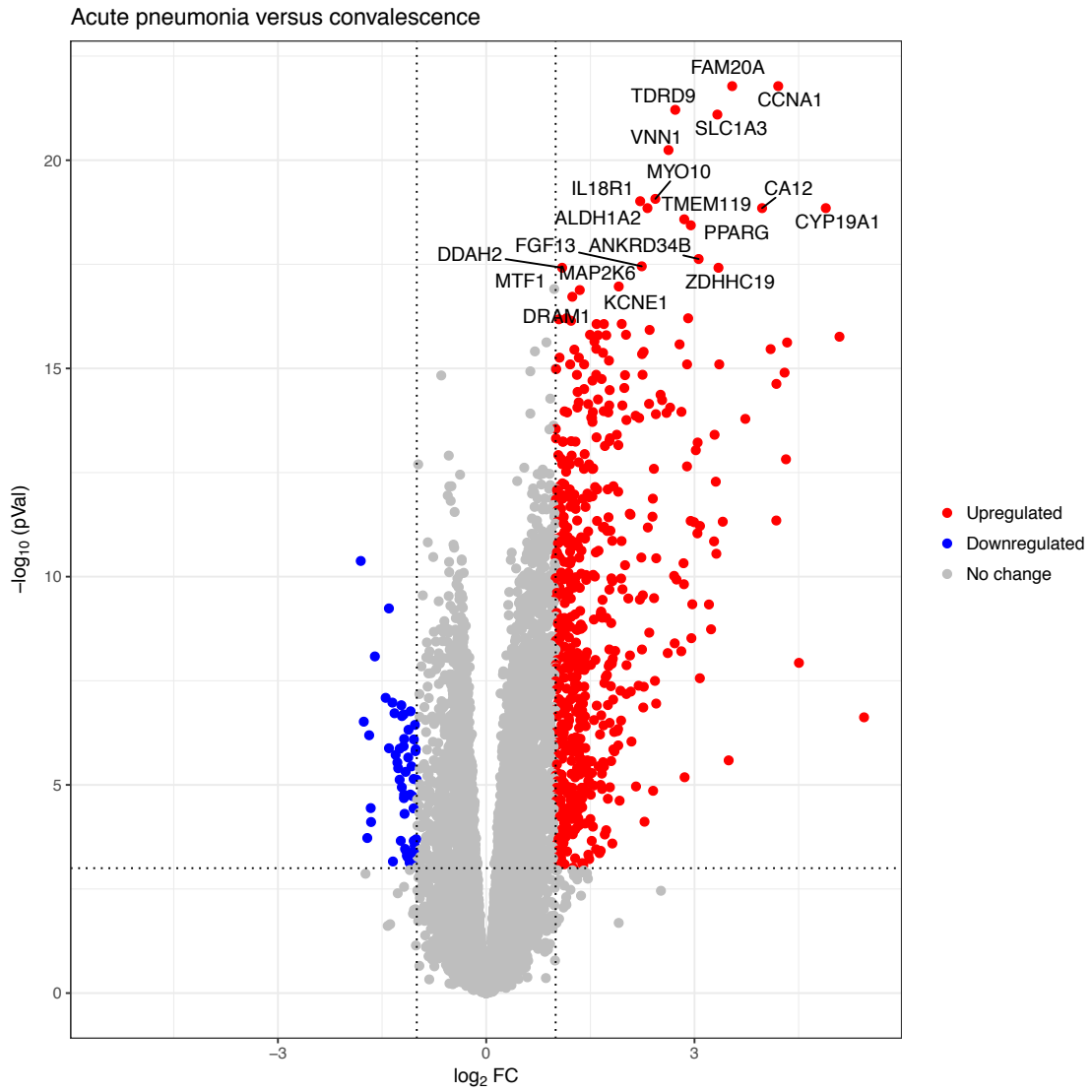


Figure 6.9: Volcano plot of differential gene expression between children with acute pneumonia (any comparator standard) and a paired convalescent sample (6–8 weeks following enrolment; FDR = 0.001). The top most differentially regulated genes are labelled with corresponding gene names.

pneumonia vs healthy controls (comparison 1, Figure 6.2), neutrophil degranulation was the most enriched pathway in children with acute pneumonia vs convalescence (comparison 2) – typically associated with an immune response to bacterial infection. Interferon α/β signalling was the second most enriched pathway – typically associated with an immune response to viral infection. Other highly enriched pathways included upregulation of the innate immune response, and downregulation of post-translational protein phosphorylation and collagen biosynthesis.

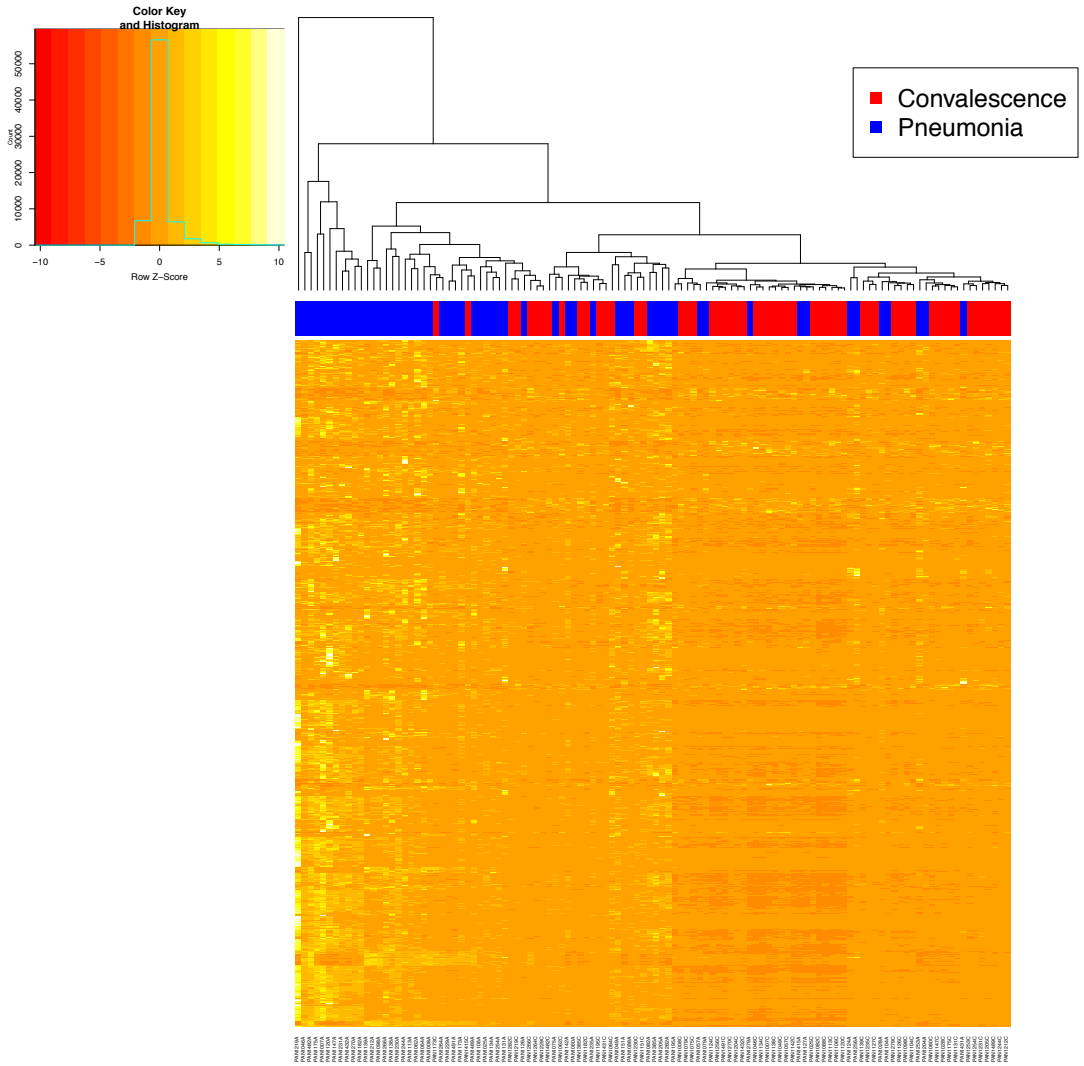


Figure 6.10: Heat map of all differentially expressed genes in children with acute pneumonia (of any comparator standard) and a paired convalescent sample (6–8 weeks following enrolment). Clustering of patients was unsupervised, and showed an increasing proportion of children with acute pneumonia towards the left of the plot.

In a subsequent analysis, I limited the comparison of differential gene expression to samples from children with acute pneumococcal pneumonia, and paired convalescent samples (n = 4). Analysis of differential gene expression between acute and convalescent samples from children with pneumococcal pneumonia showed 422 genes to be differentially expressed: 370 genes were upregulated, and 52 genes were downregulated (FDR = 0.001). The differential expression of these genes is summarised in Figure 6.12, with the top most genes annotated. The acute and convalescent samples separated into

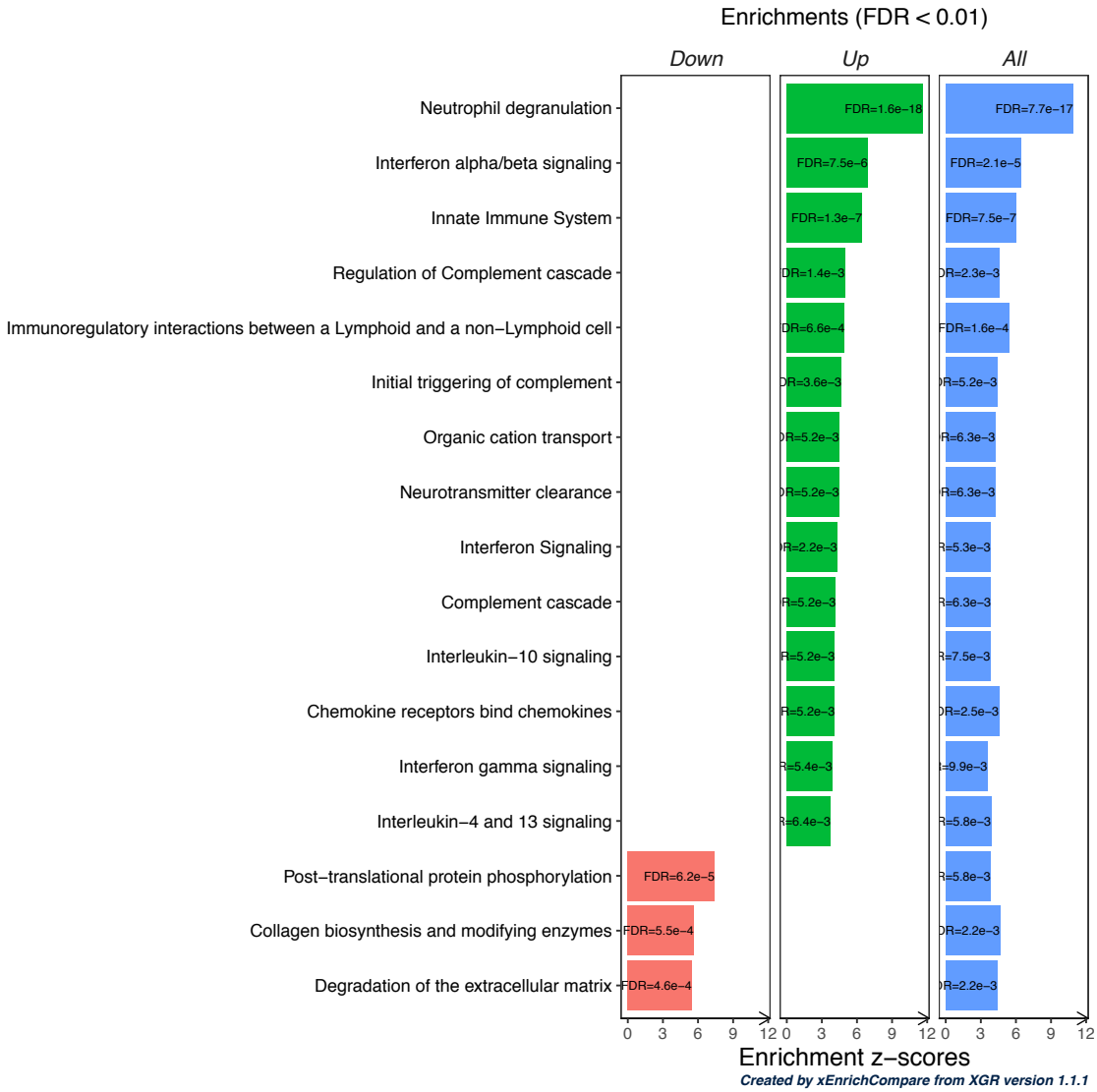


Figure 6.11: Enrichment of gene networks between children with acute pneumonia (any comparator standard) and a paired convalescent sample (6–8 weeks following enrolment).

separate clusters in unsupervised hierarchical clustering, with the partial exception of the acute sample from a single patient, PAN1415 (Figure 6.13, suffix "A"), for which the dendrogram classified as close to convalescent samples. Of note, this child (PAN1415) had a low blood neutrophil count (1.9 x 10^3/mm^3) and a low blood lymphocyte count (0.4 x 10^3/mm^3). Details for this child, and other children with pneumococcal pneumonia (and *definite other bacterial* pneumonia) are shown in Table 6.2. As with previous comparisons, pathways associated with immunological activation – neutrophil degranulation, interferon α/β signalling, innate immunity, the complement cascade and immunoregulation were all enriched and upregulated (Figure 6.11).

Table 6.2: Characteristics of children with *definite pneumococcal* (DP), *probable pneumococcal* (PP) and *definite other bacterial* (DOB) pneumonia on whom transcriptomic investigations were undertaken, by pneumonia aetiology comparator standard. EPC, endpoint consolidation; ST, serotype.

Child	**PAN1007**	**PAN1041**	**PAN1048**	**PAN1078**	**PAN1102**	**PAN1120**	**PAN1136**
Comparator standard	DP	DP	PP	DP	DOB	DOB	DOB
Sex	M	M	M	F	M	F	F
Age (years)	9.8	5.7	11.3	2.9	0.6	0.6	5.2
Length of illness (days)	4	6	6	5	6	4	2
CRP mg/l	250	317	212	276	14	206	88
Neutrophil count x 10^3/mm^3	20	11	16	3	9	11	-
Lymphocyte count x 10^3/mm^3	2.2	2.7	1.1	1.6	8.8	2.6	-
Radiograph result	EPC	EPC	EPC	EPC	EPC	Normal	-
Pneumococcal carriage serotype	1	None	1	19A	None	39	None
Prior antibiotics	N	N	N	Y	Y	Y	N
Culture result	*S. pneumoniae*	*S. pneumoniae*	Negative	*S. pneumoniae*	*Acinetobacter* spp.	*S. enterica* var Typhi	*S. enterica* var Typhi
Child	**PAN1175**	**PAN1219**	**PAN1222**	**PAN1239**	**PAN1244**	**PAN1385**	
Comparator standard	PP	DP	DOB	PP	DOB	PP	
Sex	M	M	M	M	M	M	
Age (years)	8.6	9.1	9	8.2	0.7	6.8	
Length of illness (days)	2	2	6	4	1	3	
CRP mg/l	158	102	185	207	15	66	
Neutrophil count x 10^3/mm^3	28	19	10	29	15	13	
Lymphocyte count x 10^3/mm^3	2.1	1.7	0.8	4.1	3.9	3.3	
Radiograph result	EPC	EPC	EPC	EPC	Infiltrates	Normal	
Pneumococcal carriage serotype	1	None	None	1	None	1	
Prior antibiotics	N	Y	Y	Y	Y	N	
Culture result	Negative	*S. pneumoniae*	*S. aureus*	Negative	*S. aureus*	Negative	
Child	**PAN1415**	**PAN1432**	**PAN1434**	**PAN1458**	**PAN1483**	**PAN1491**	
Comparator standard	DP	DP	DP	DOB	DOB	DOB	
Sex	M	M	M	F	F	M	
Age (years)	3.8	5.8	7.8	0.7	7.6	0.9	
Length of illness (days)	3	6	1	7	8	4	
CRP mg/l	49	58	184	2	6	13	
Neutrophil count x 10^3/mm^3	1.9	15	9.5	1.8	9.7	4.3	
Lymphocyte count x 10^3/mm^3	0.4	5	1.8	9.8	9.3	8.1	
Radiograph result	Normal	EPC	EPC	Normal	Infiltrates	Normal	
Pneumococcal carriage ST	None	6B	None	None	None	10A	
Prior antibiotics	Y	Y	N	Y	N	Y	
Culture result	*S. pneumoniae*	*S. pneumoniae*	*S. pneumoniae*	*E. coli*	*Acinetobacter* spp.	*N. meningitidis*	

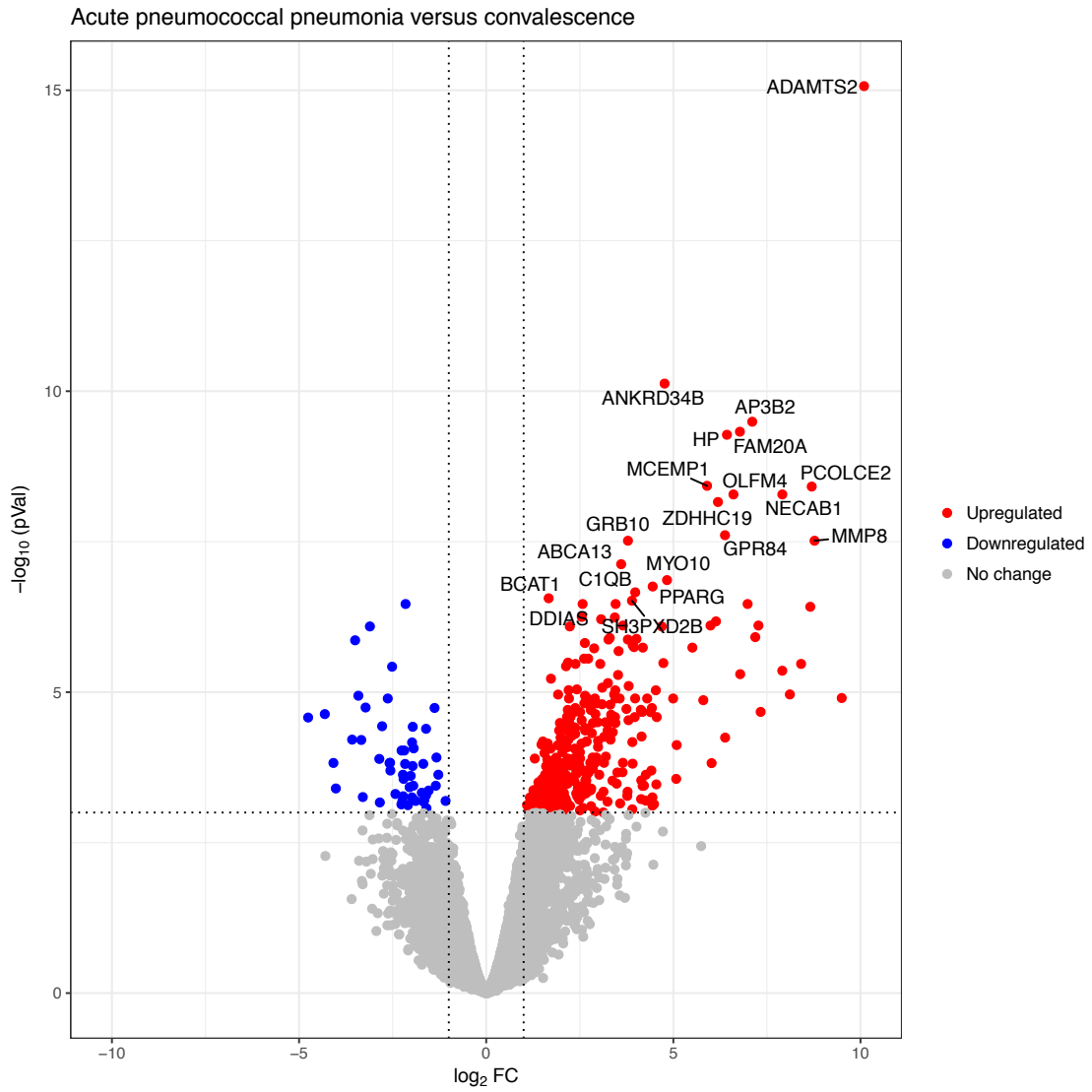


Figure 6.12: Volcano plot of differential gene expression between children with pneumococcal pneumonia and a paired convalescent sample (6–8 weeks following enrolment; FDR = 0.001). The top most differentially regulated genes are labelled with corresponding gene names.

6.3.5 Comparison 3: bacterial pneumonia vs viral pneumonia

I examined differential gene expression between children with bacterial pneumonia (i.e. *definite pneumococcal*, *probable pneumococcal* and *definite other bacterial* pneumonia: total n = 19) vs viral pneumonia (i.e. *probable viral* pneumonia; n = 25). The analysis showed 1262 genes to be differentially expressed: 1097 genes were upregulated, and 165 genes were downregulated (FDR = 0.001). The differential expression of these genes is summarised in Figure 6.15, with the top most genes annotated. Unsupervised clustering

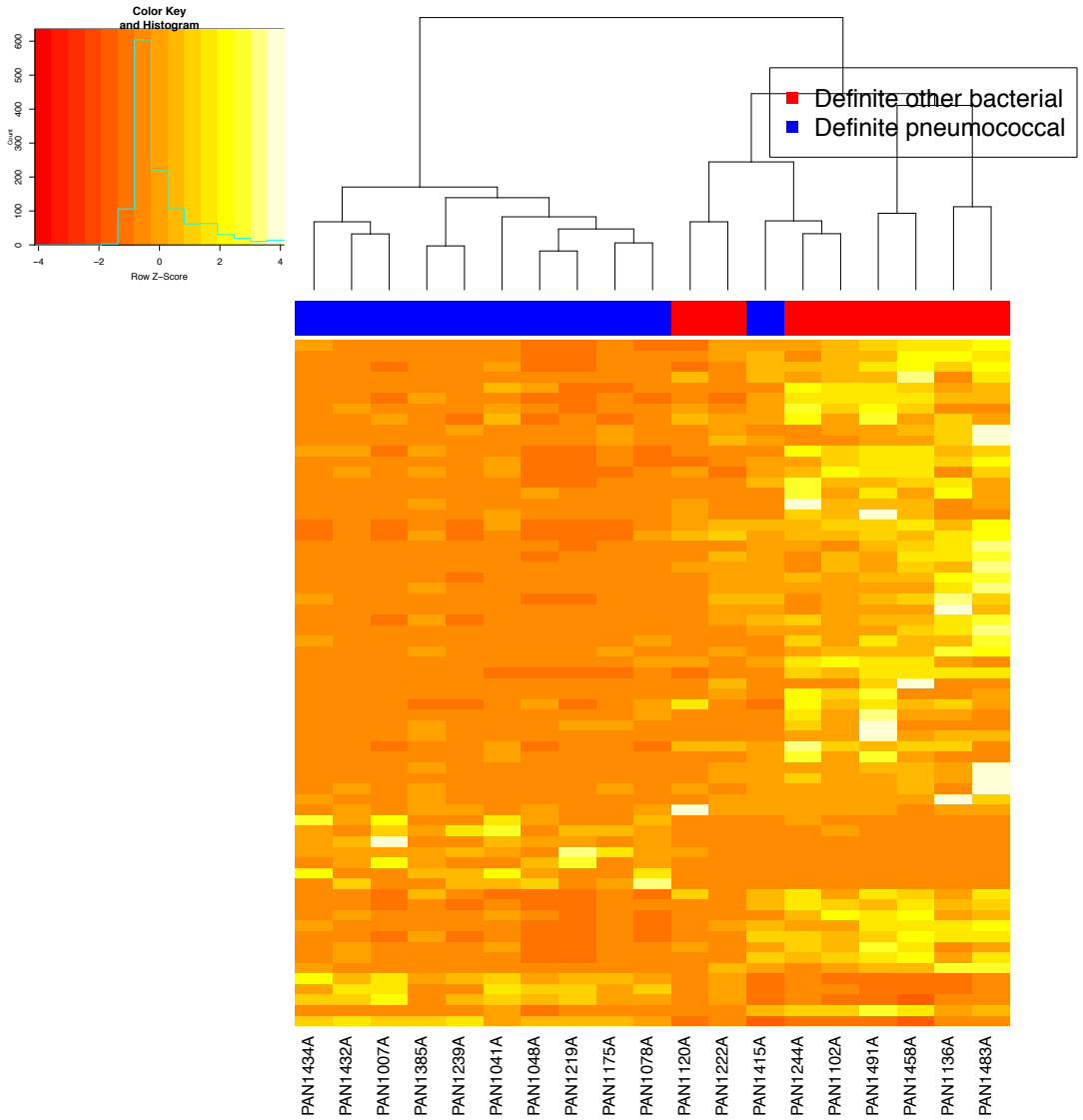


Figure 6.13: Heat map of all differentially expressed genes in children with pneumococcal pneumonia and a paired convalescent sample (6–8 weeks following enrolment). Clustering of patients was unsupervised, and showed an increasing proportion of children with acute pneumonia towards the left of the plot.

showed 12/19 samples from children with bacterial pneumonia clustering together with an early split on the dendrogram from other samples (Figure 6.16). However, the remaining 7/19 samples from children with bacterial pneumonia did not cluster together in this analysis. Examination of the clinical characteristics of these seven children (Table 6.2) showed only one child to have a plasma CRP concentration >60 mg/l and/or a neutrophil count >12 x 10^3/mm^3 (PAN1385, 66 mg/l and 13 x 10^3/mm^3, respectively).

As with previous comparisons, pathways associated with immunological activation were

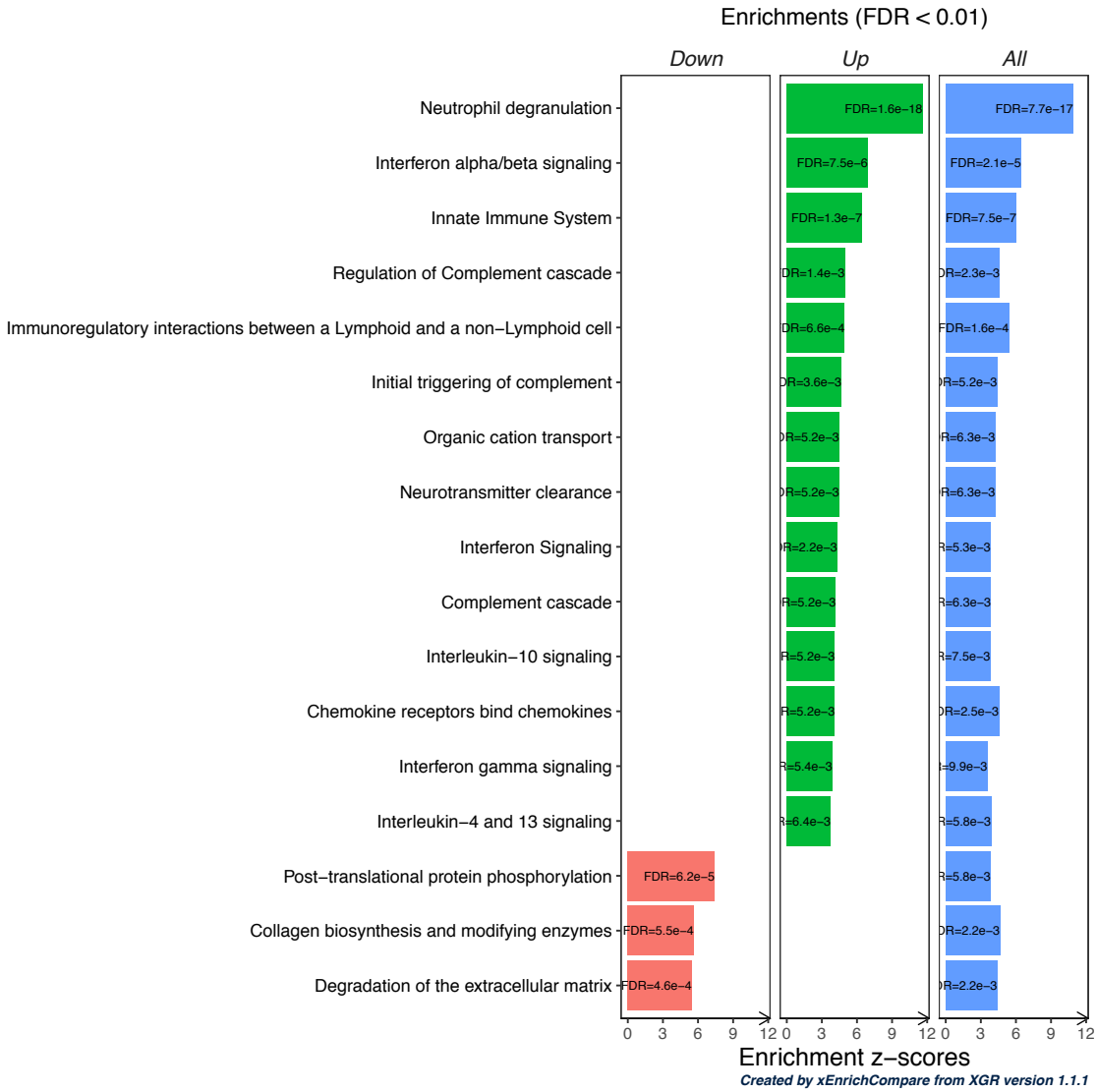


Figure 6.14: Enrichment of gene networks between children with pneumococcal pneumonia and a paired convalescent sample (6–8 weeks following enrolment).

enriched and upregulated in bacterial pneumonia (Figure 6.17). Pathways more typically associated with bacterial infection – neutrophil degranulation, and toll-like receptor signalling for example – were clearly enriched, in contrast to pathways more typically associated with viral infection or tuberculosis – interferon α/β signalling – which were not enriched. (However, interferon α/β signalling was also upregulated in the comparison of acute pneumococcal pneumonia vs convalescence.) Pathways associated with oxygen and carbon dioxide exchange in erythrocytes were enriched, and downregulated.

Using the genes identified in comparison 3, a "Bacterial Gene Score" to discriminate

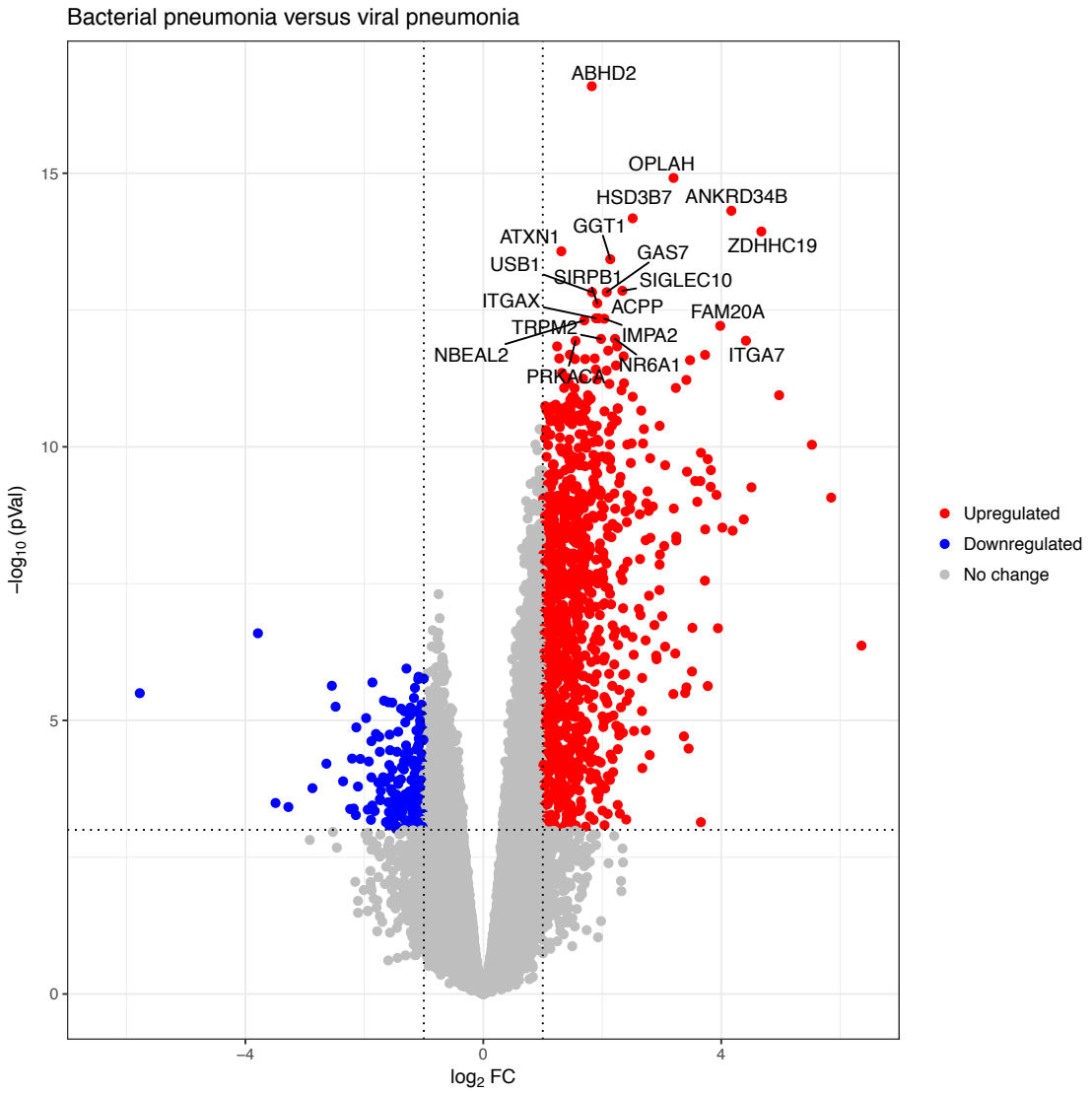


Figure 6.15: Volcano plot of differential gene expression between children with bacterial pneumonia and viral pneumonia (FDR = 0.001). The top most differentially regulated genes are labelled with corresponding gene names.

between bacterial and viral pneumonia was calculated by a simple summation the vectors of differentially expressed genes (FDR = 0.001) to create a single value for each sample (see Methods). The Bacterial Gene Scores for each sample are plotted in Figure 6.18, including the corresponding bacterial isolate. There was a highly significant difference in gene scores between the two groups (two sample t-test, p = 1.088 x 10^{-10}). ROC curve analysis of these results produced a single optimal threshold of the Bacterial Gene Score to optimise the area under the curve, giving a sensitivity of 0.68 and specificity of 1.0 (Figure 6.19). Application of this threshold to samples from children in all pneumonia

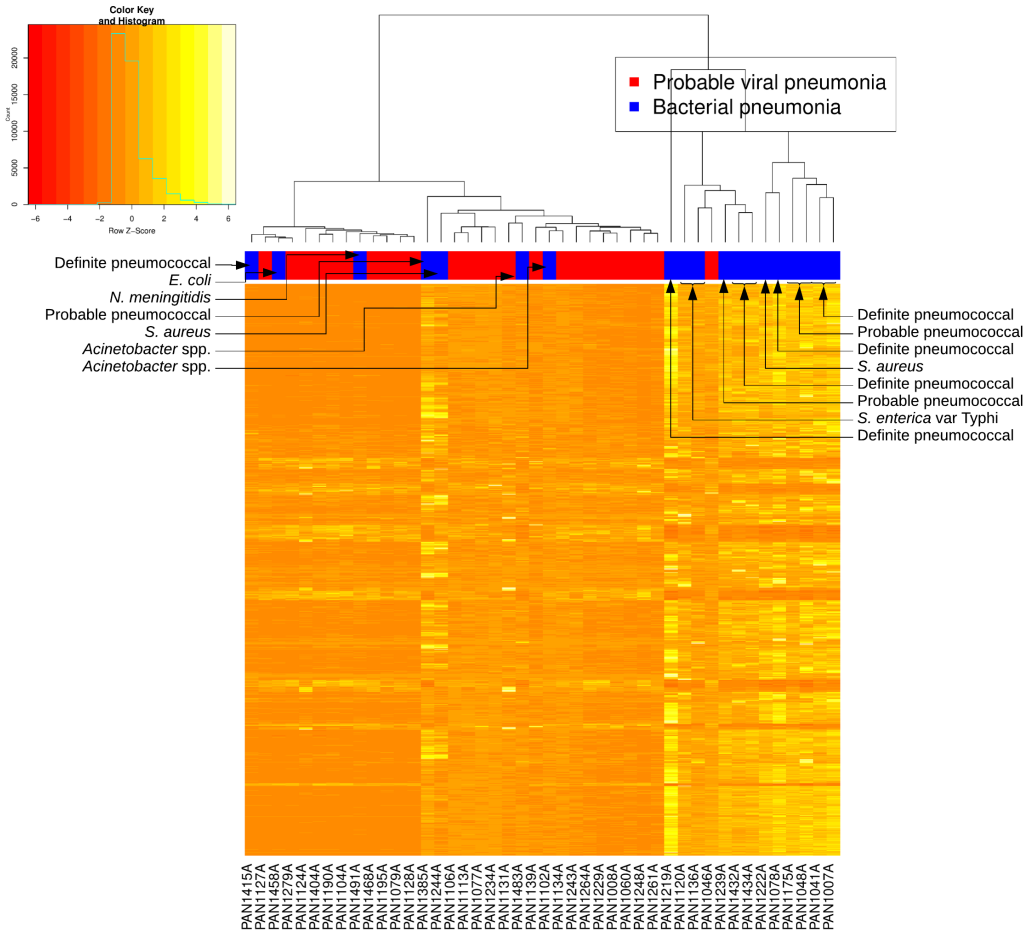


Figure 6.16: Heat map of all differentially expressed genes in children with bacterial pneumonia vs viral pneumonia. Unsupervised analyses show a clear cluster of samples from children with bacterial pneumonia – including the majority of children with pneumococcal pneumonia – as an early branch of the dendrogram to the right of the heat map. Samples from children with pneumococcal pneumonia, or bacterial isolates from children with *definite other bacterial* pneumonia, are marked.

aetiology comparator standards and healthy controls is shown in Figure 6.20.

The number of children with bacterial pneumonia and viral pneumonia were too few for the generation of a "discovery" and "validation" group to test the Bacterial Gene Score. I therefore used "leave-one-out" cross validation to classify samples according to their Bacterial Gene Score into "bacterial" and "viral" groups. This technique creates a series of datasets in an iterative manner by sequentially removing one of the samples (i) from the dataset, recalculating the diagnostic threshold, and reapplying it to the dataset $n - i$. The classification of samples to "bacterial" or "viral" pneumonia according to the gene

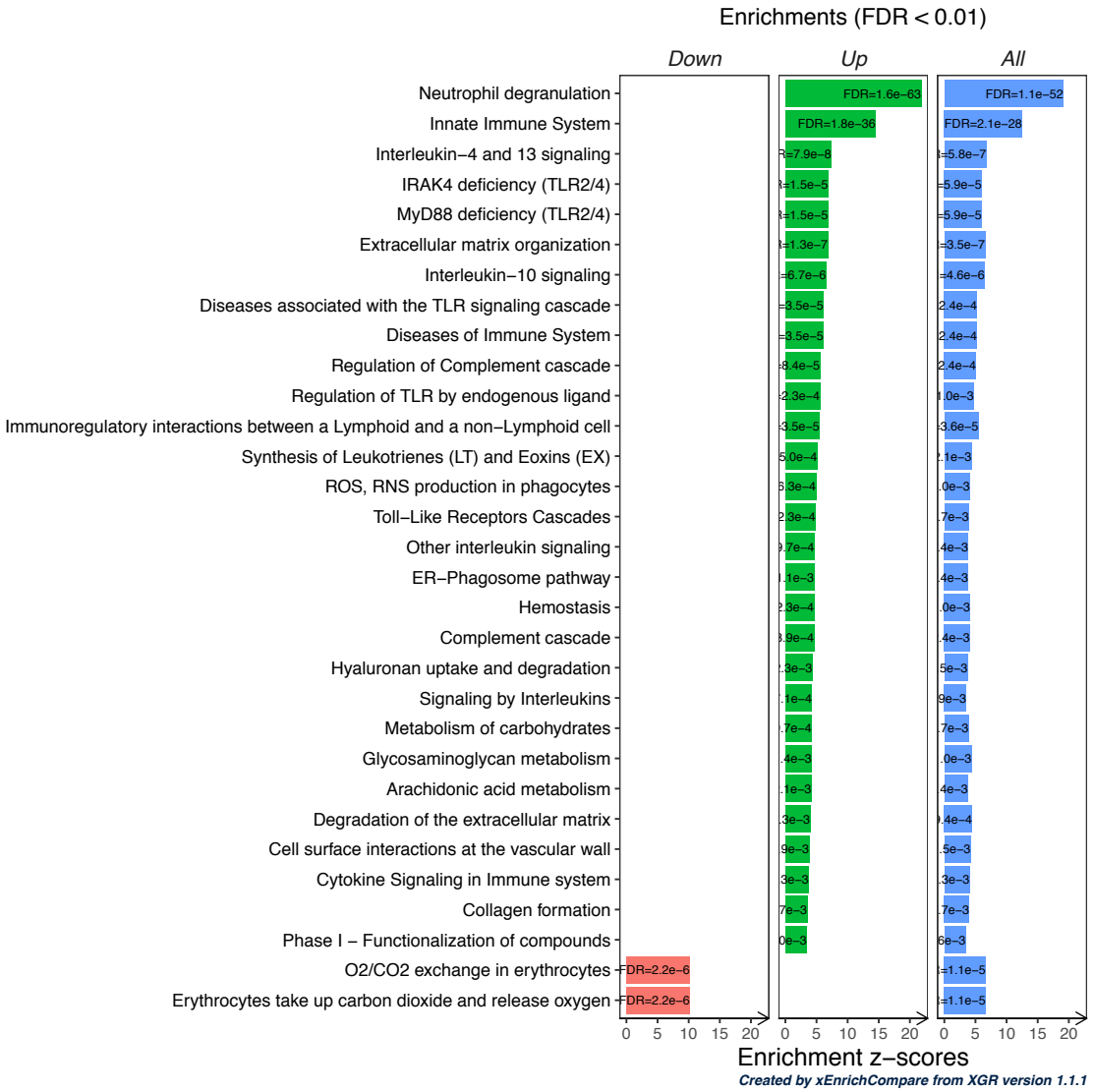


Figure 6.17: Enrichment of gene networks between children with bacterial pneumonia and viral pneumonia. Networks were derived from the REACTOME ontology using the XGR package in R as detailed in the Methods.

score is detailed in Table 6.3. The Bacterial Gene Score performed poorly at classifying samples from children with *definite other bacterial* pneumonia (sensitivity 2/6), but well at classifying samples from children with pneumococcal pneumonia (sensitivity 10/11), as "bacterial". In addition all samples from children with *probable viral* pneumonia and healthy controls were classified as "viral" (in the case of healthy controls this can be inferred to mean "non-bacterial"). 21/40 samples from children with *probable bacterial* pneumonia were classified as "bacterial", and 5/26 children with *unknown* aetiology were classified as "bacterial".

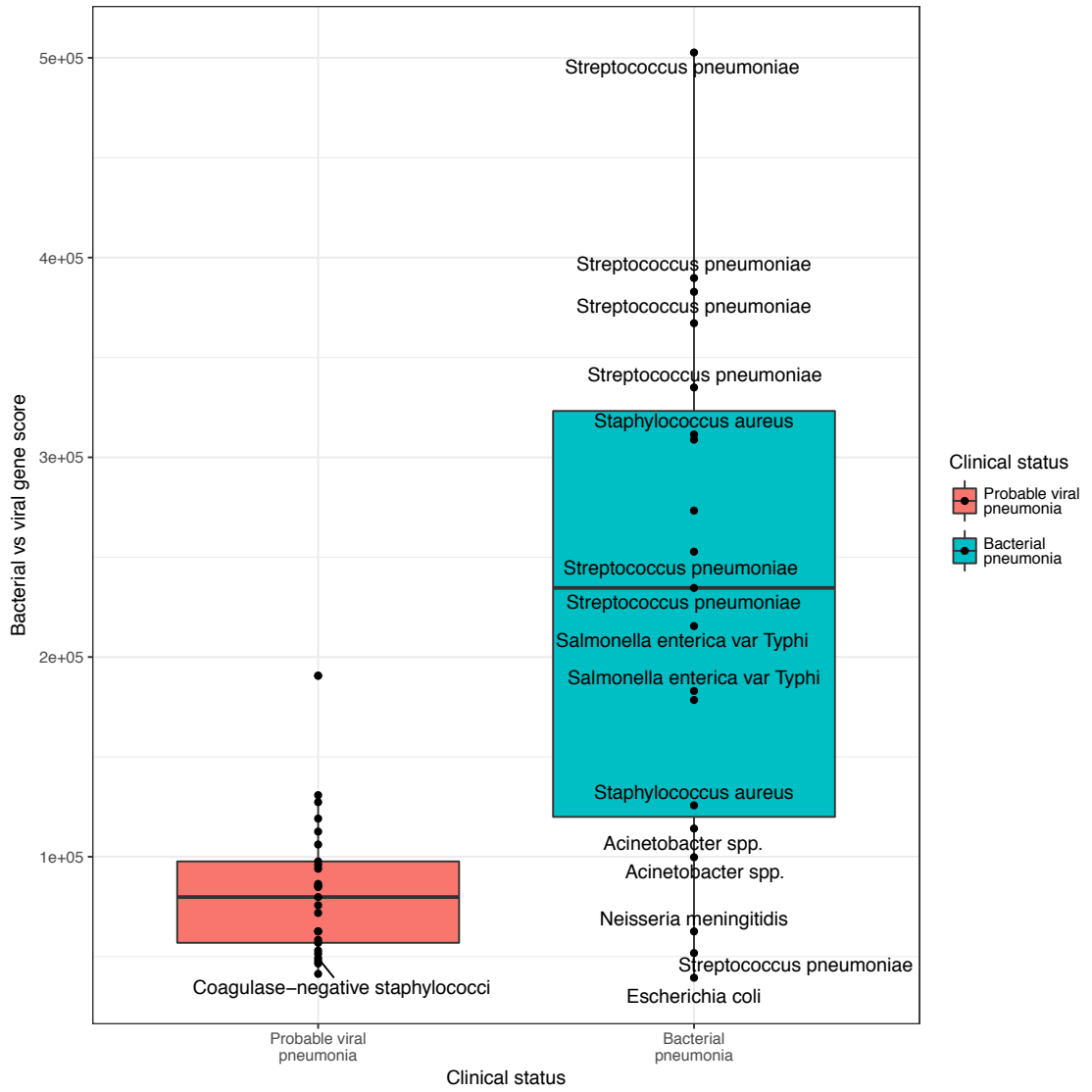


Figure 6.18: Application of a transcriptomic signature to discriminate between bacterial pneumonia and viral pneumonia, the Bacterial Gene Score ($p = 1 \times 10^{-10}$). Each point represents a single child with pneumonia. Children with positive blood or pleural fluid culture results are labelled.

6.3.6 Comparison 4: pneumococcal pneumonia vs other bacterial pneumonia

I examined differential gene expression between children with pneumococcal pneumonia (i.e. *definite pneumococcal* and *probable pneumococcal* pneumonia; n = 11) and *definite other bacterial* pneumonia (n = 8). Due to the small number of genes differentially expressed at an FDR of 0.001, I used a less stringent threshold of FDR = 0.01. The analysis showed 65 genes to be differentially expressed: 11 genes were upregulated in

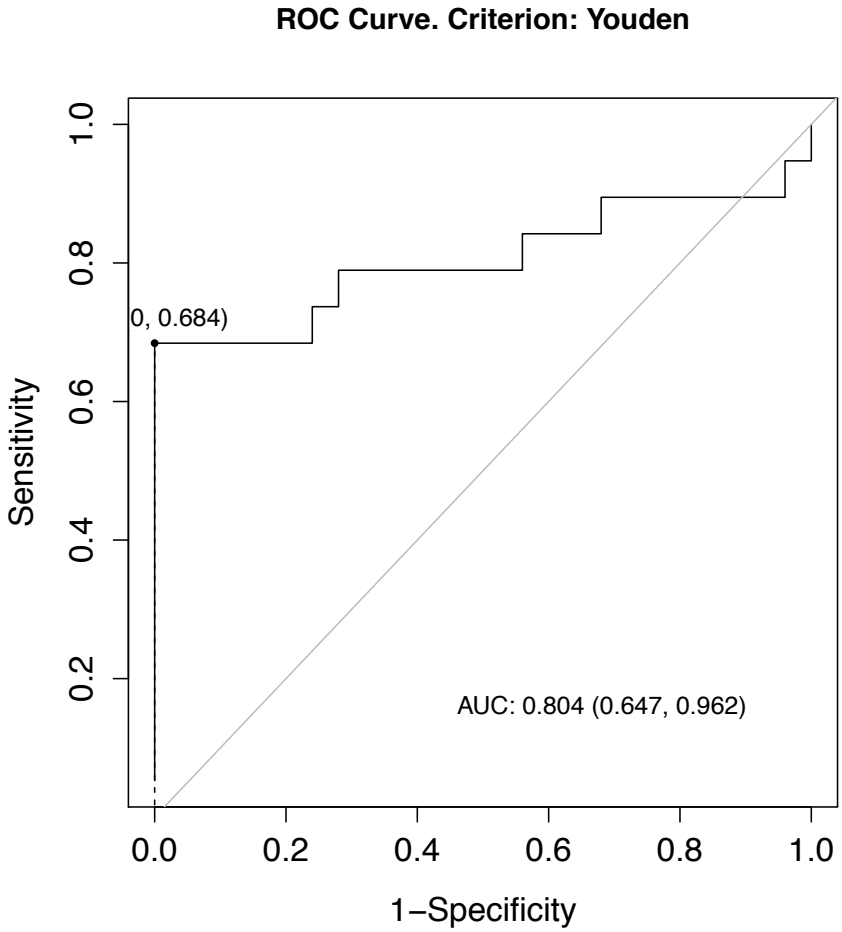


Figure 6.19: ROC curve of a transcriptomic signature to discriminate between bacterial pneumonia and viral pneumonia, the Bacterial Gene Score. The optimal threshold (as determined by the Youden Index) gave a sensitivity of 0.68 (950.43, 0.87) and a specificity of 1.0 (95% CI 0.86, 1.00) and an area under the curve (AUC) of 0.82.

pneumococcal pneumonia, and 54 genes were downregulated in comparison to *definite other bacterial* pneumonia. Figure 6.21 shows this differential gene expression, with the top most genes annotated. Unsupervised clustering showed 10/11 samples from children with pneumococcal pneumonia clustering together, with one outlier (from child PAN1415, as discussed earlier, Figure 6.22). Pathway analysis showed enrichment of "immunoregulatory interactions between a lymphoid and a non-lymphoid cell", which was downregulated in pneumococcal pneumonia (FDR = 0.01).

I used a similar approach to that detailed for comparison 3 (above) to develop a gene score to discriminate between pneumococcal and *definite other bacterial* pneumonia (the "Pneumococcal Gene Score") as a single value for each sample. The gene scores

Table 6.3: Diagnostic accuracy of the Bacterial Gene Score, using a threshold derived from ROC curve analysis of bacterial vs viral pneumonia. The numbers of children classified as "viral" or "bacterial" by the application of the optimal threshold of the bacterial vs viral gene score in pneumonia aetiology comparator standards is described in the lower half of the table.

	Bacterial Gene Score	
Classical diagnostic accuracy		
Sensitivity (95% CI)	0.68 (0.43, 0.87)	
Specificity (95% CI)	1.0 (0.86, 1.0)	
Positive predictive value	1	
Negative predictive value	0.81	
"Leave-one-out" validation	**"Viral"**	**"Bacterial"**
Definite pneumococcal	1	6
Definite other bacterial	6	2
Probable pneumococcal	0	4
Probable bacterial	19	21
Unknown	21	5
Probable viral	25	0
Healthy control	25	0

for each sample are plotted in Figure 6.23, including the corresponding bacterial isolate. There was a highly significant difference in gene scores between the two groups (Kruskal-Wallis rank sum test, $p < 0.001$). ROC curve analysis produced a single optimal threshold, giving a sensitivity of 0.91 and a specificity of 1.0 (Figure 6.24). Application of this threshold to samples from children in all pneumonia aetiology comparator standards and healthy controls is shown in Figure 6.25.

I used the Pneumococcal Gene Score and cross validation to classify samples from children in all pneumonia aetiology comparator standards (and healthy controls) into two groups: "pneumococcal pneumonia" and "non-pneumococcal pneumonia". The classification of samples to "pneumococcal" or "non-pneumococcal" pneumonia according to the gene score is detailed in Table 6.4. The Pneumococcal Gene Score performed well at classifying samples from children with pneumococcal pneumonia (sensitivity 9/11) vs *definite other bacterial* pneumonia (0/8) as "pneumococcal pneumonia". All children with *probable viral* pneumonia, and all healthy controls were classified as "non-pneumococcal" pneumonia. In addition, 3/40 samples from children with *probable bacterial* pneumonia, and 2/24 samples from children with *unknown* aetiology were classified as "pneumococcal pneumonia" using this threshold. Small variations in the

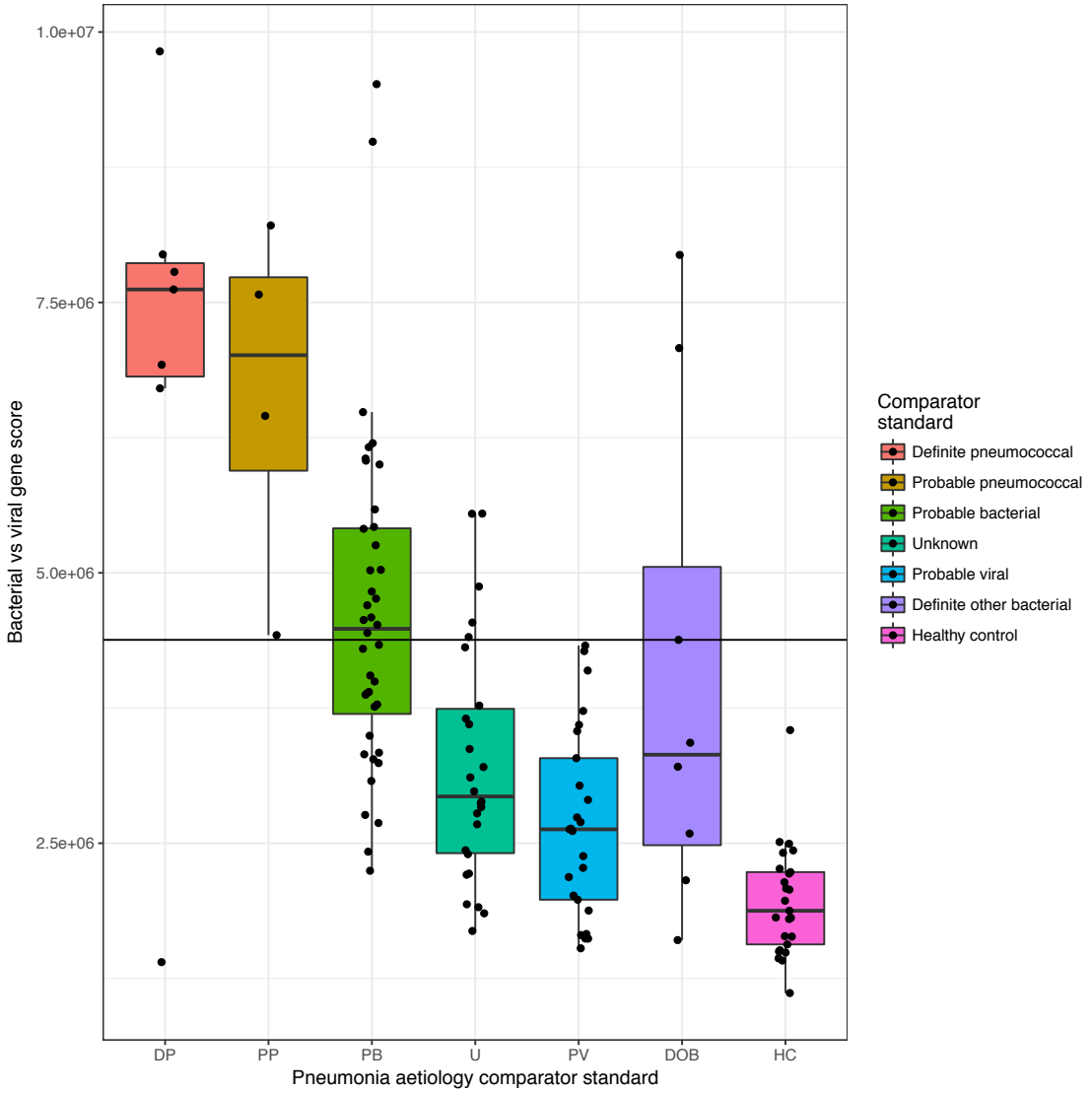


Figure 6.20: Boxplots showing the Bacterial Gene Score applied to children with pneumonia of all comparator standards and healthy controls. The optimal threshold (as determined by the Youden Index) was determined from ROC curve analysis of bacterial vs viral pneumonia. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control.

threshold of the Pneumococcal Gene Score produced considerable variability in the number of samples that were classified as "pneumococcal pneumonia".

10 genes were differentially expressed in both a comparison between acute pneumococcal pneumonia vs convalescence (see above), and in this comparison between pneumococcal pneumonia and *definite other bacterial* pneumonia. These genes are detailed in Table 6.5.

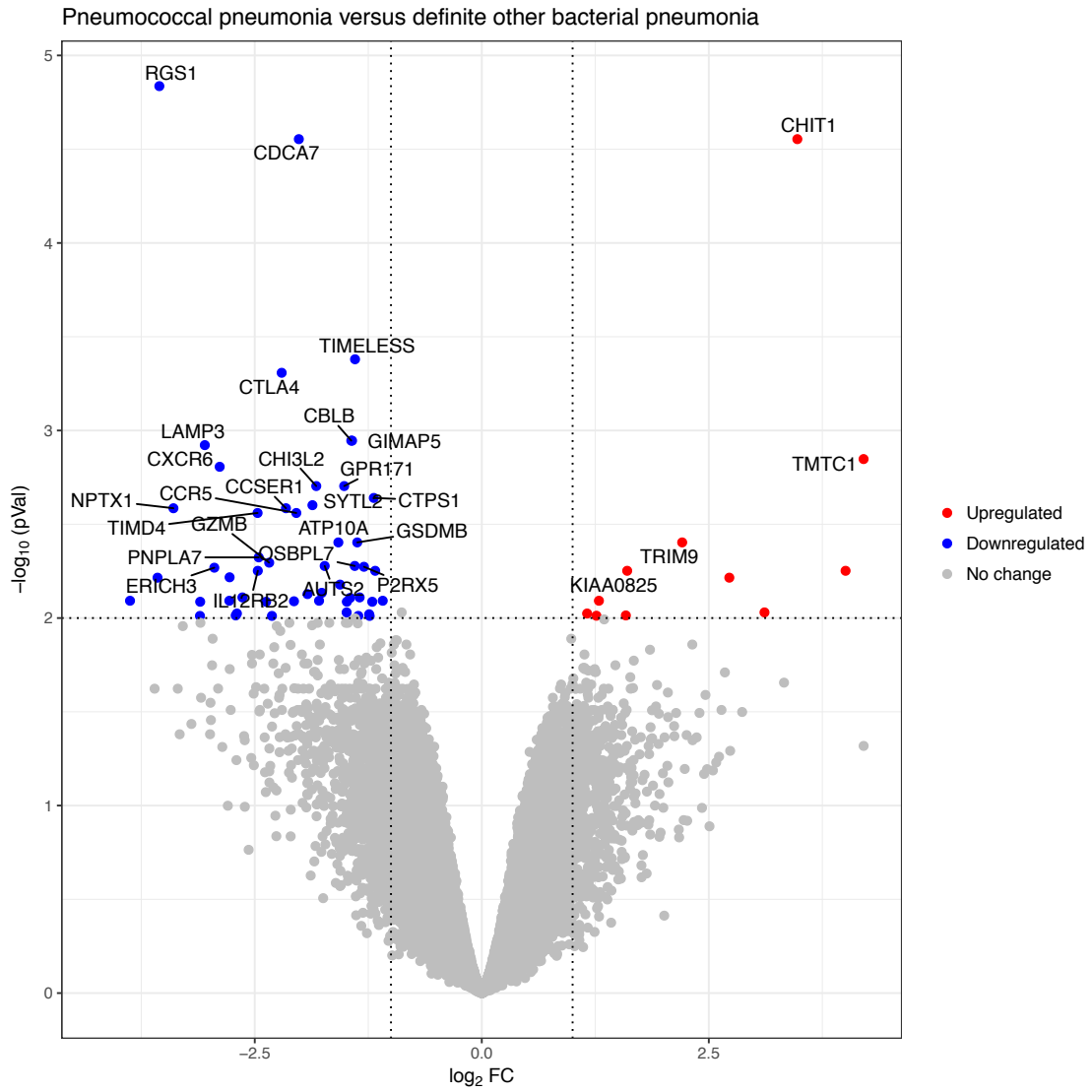


Figure 6.21: Volcano plot of differential gene expression between children with pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score (FDR = 0.01). The top most differentially regulated genes are labelled with corresponding gene names.

6.4 Discussion

The work presented in this chapter describes the transcriptome of children with pneumonia in Nepal. Comparisons between children with pneumonia classified according to an *a priori* comparator standard of pneumonia aetiology were used to explore differential gene expression between different likely aetiologies of pneumonia. I also described the relationship between acute childhood pneumonia and convalescence. I then formally assessed the ability a gene scores for bacterial pneumonia and a gene score for

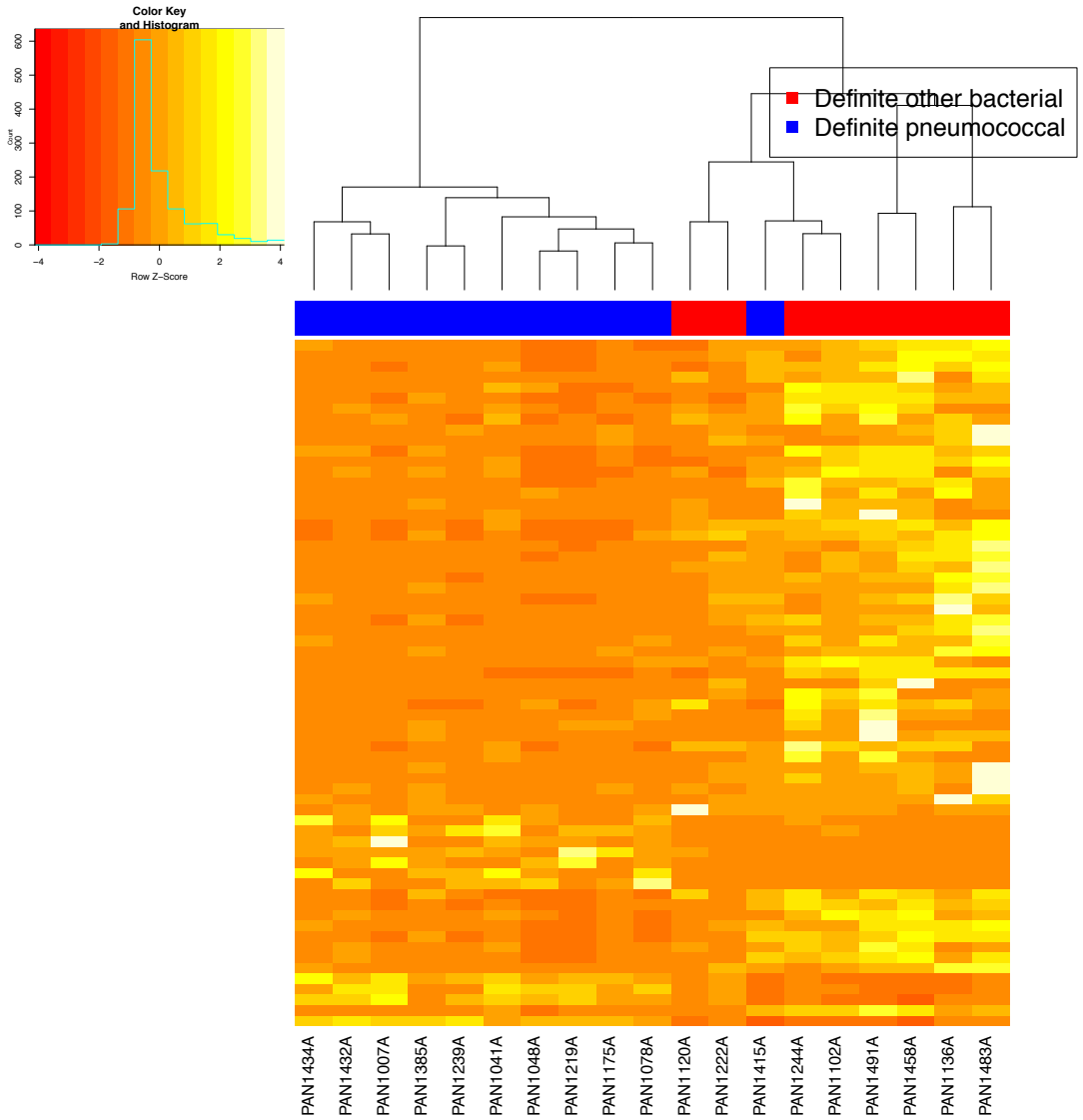


Figure 6.22: Heat map of all differentially expressed genes in children with pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score. Clustering of patients was unsupervised, and showed an increasing proportion of children with bacterial pneumonia towards the left of the plot.

pneumococcal pneumonia to discriminate between pneumonia aetiologies.

Here, I used RNA sequencing to examine gene expression levels, providing a "digital" count of expressed transcripts. Most previous analyses of differential gene expression have used microarray technology, which used fluorescence intensity of probes to RNA transcripts to quantify gene expression. This makes formal integration of datasets bioinformatically challenging. Nonetheless, comparisons between results from various groups are valid and relevant.

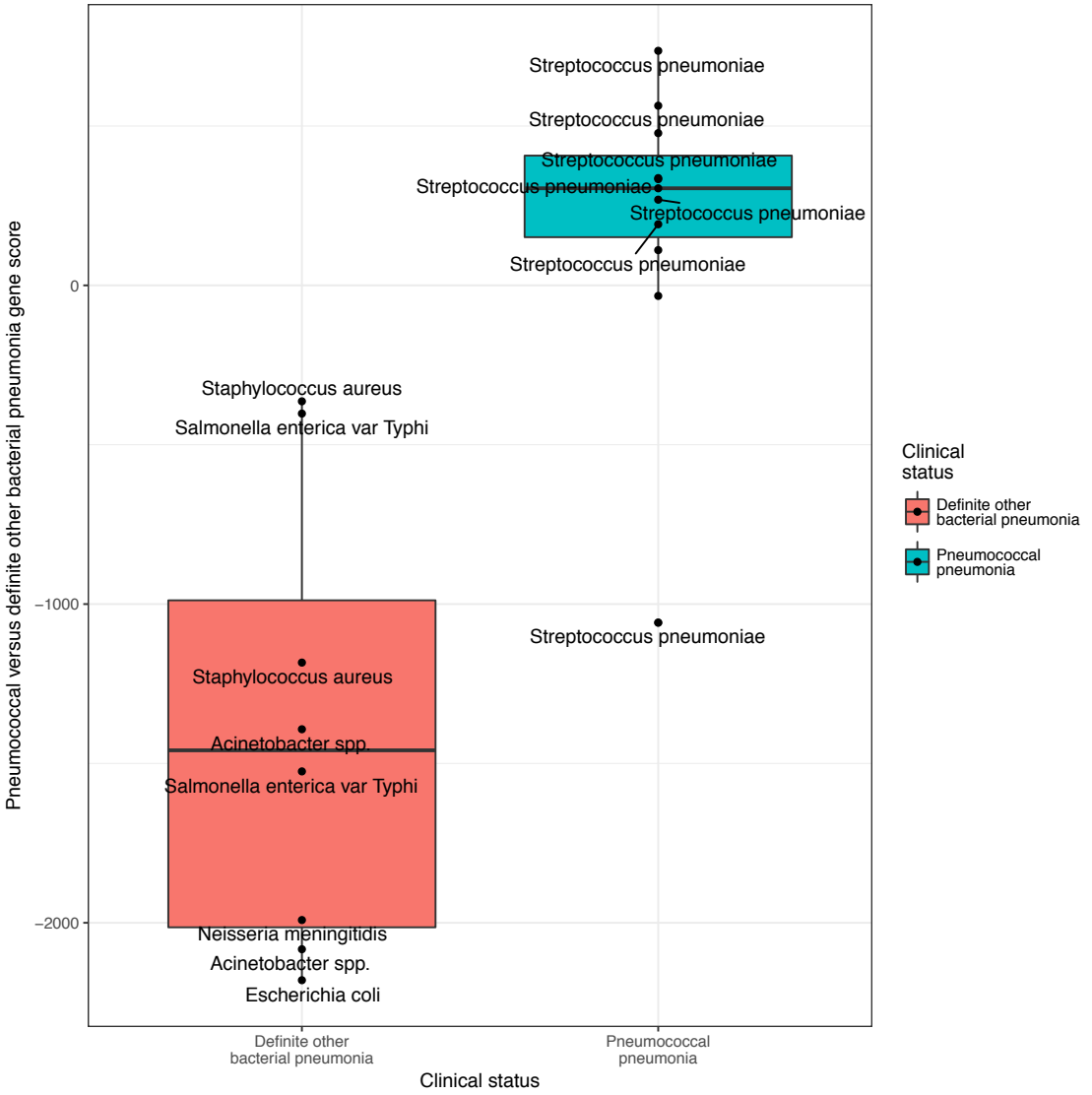


Figure 6.23: Boxplots describing a transcriptomic signature to discriminate between pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score ($p < 0.001$). Each point represents a single child with pneumonia. Children with positive blood or pleural fluid culture results are labelled.

6.4.1 The transcriptome discriminates between healthy controls and disease

Principal component analyses of differential gene expression between children with pneumonia and healthy controls showed overlapping clusters of samples (Figure 6.4) on principal components 1 and 2. However, samples from children with pneumococcal pneumonia were largely distinct from samples from healthy controls and children with *probable viral* pneumonia. Samples from children with *unknown* aetiology and *probable*

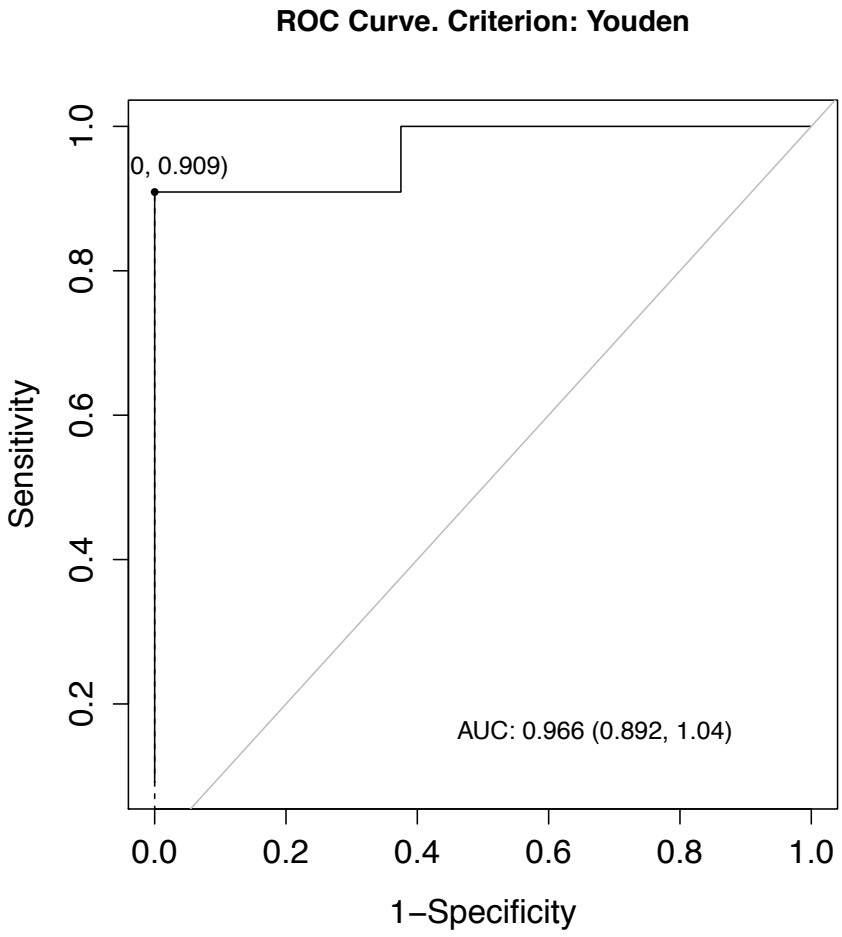


Figure 6.24: ROC curve of a transcriptomic signature to discriminate between pneumococcal pneumonia and *definite other bacterial* pneumonia, the Pneumococcal Gene Score. The optimal threshold (as determined by the Youden Index) gave a sensitivity of 0.91 and a specificity of 1.0.

bacterial pneumonia were intermediate. Samples from children with *definite other bacterial* pneumonia were distinct from each other, and mainly distinct from samples from children with *probable viral* pneumonia and healthy controls. This suggests a model whereby the transcriptome of children with pneumococcal pneumonia, and some children with *probable bacterial* pneumonia, "moves" across the principal components during acute illness.

The concept of the transcriptome in disease moving away from a baseline (defined by the transcriptome of healthy controls) has been formalised as the "molecular distance to health"[192, 279]. In this method, the "molecular distance" of a sample is calculated by comparing the expression levels of a single gene in a patient with disease to the mean

Table 6.4: Diagnostic accuracy of the Pneumococcal Gene Score, using a threshold derived from ROC curve analysis of pneumococcal vs *definite other bacterial* gene score. The numbers of children classified as "non-pneumococcal" or "pneumococcal" by the application of the optimal threshold of the Pneumococcal Gene Score in pneumonia aetiology comparator standards is described in the lower half of the table.

	Pneumococcal Gene Score	
Classical diagnostic accuracy		
Sensitivity (95% CI)	0.91 (0.59, 1.0)	
Specificity (95% CI)	1.0 (0.63, 1.0)	
Positive predictive value	1	
Negative predictive value	0.89	
"Leave-one-out"" validation"	**"Non-pneumococcal"**	**"Pneumococcal"**
Definite pneumococcal	1	6
Definite other bacterial	8	0
Probable bacterial	37	3
Probable pneumococcal	1	3
Unknown	24	2
Probable viral	25	0
Healthy controls	25	0

and variance of the expression value of this gene in the healthy control group. Non-differentially expressed genes are then filtered. The sum of differential gene expression the forms the molecular distance to health. In separate cohorts of children with respiratory viral infections the molecular distance to health was able to predict the severity of illness in children with respiratory syncytial virus (RSV) infection[200]. Separate analyses have developed this concept further to examine the host immune response to NP carriage of potential respiratory pathogens in healthy children, children with mild infection and children with severe infection[280], showing that distinct clusters of differential gene expression are formed in response to the carriage of specific pathogens. This approach may be of considerable use in studies of pneumonia aetiology – where differentiating healthy carriage from infection has proved difficult (see Chapter 1).

Here, using paired acute and convalescent samples, I showed that the transcriptome returns to a tightly clustered focus in convalescence – regardless of the aetiology of the acute pneumonia (Figure 6.8). However, convalescent samples do not completely overlap with healthy controls (Figure 6.4), suggesting that transcriptomic perturbations persist beyond two months. A similar finding was noted in children with RSV bronchiolitis, with blood immune profiles remaining altered from healthy controls one month following

Table 6.5: Genes differentially expressed in both a comparison of acute pneumococcal pneumonia and convalescence and in a comparison of pneumococcal pneumonia and *definite other bacterial* pneumonia.

Gene name	Full name	Function[a]	Regulation (acute pneumococcal pneumonia vs convalescence)	Regulation (pneumococcal vs definite other bacterial pneumonia)
RGS1	Regulator of G protein signalling 1	Downstream signalling of leukotriene receptors, inhibits B cell chemotaxis	Down	Down
CHIT1	Chitotriosidase-1	Degrades chitin, defence against nematodes	Up	Up
GIMAP5	GTPase IMAP family member 5	Mitochondrial integrity. Contributes to T cell quiescence	Down	Down
CXCR6	C-X-C chemokine receptor type 6	Receptor for the C-X-C chemokine CXCL16	Down	Down
SYTL2	Synaptotagmin-like protein 2	Required for cytotoxic granule exocytosis in lymphocytes	Down	Down
MCOLN2	Mucolipin-2	Calcium channel activity	Down	Down
KLRK1	NKG2-D type II integral membrane protein	Stimulatory innate immune responses to natural killer cells. T cell activator	Down	Down
CA8	Carbonic anhydrase-related protein	*unknown* (does not have carbonic anydrase catalytic activity)	Down	Down
FSTL4	Follistatin-related protein 4	Calcium binding	Up	Up
SLFN12L	Schlafen family member 12-like protein	ATP binding	Down	Down

[a]Derived from UniProt, *http://www.uniprot.org*

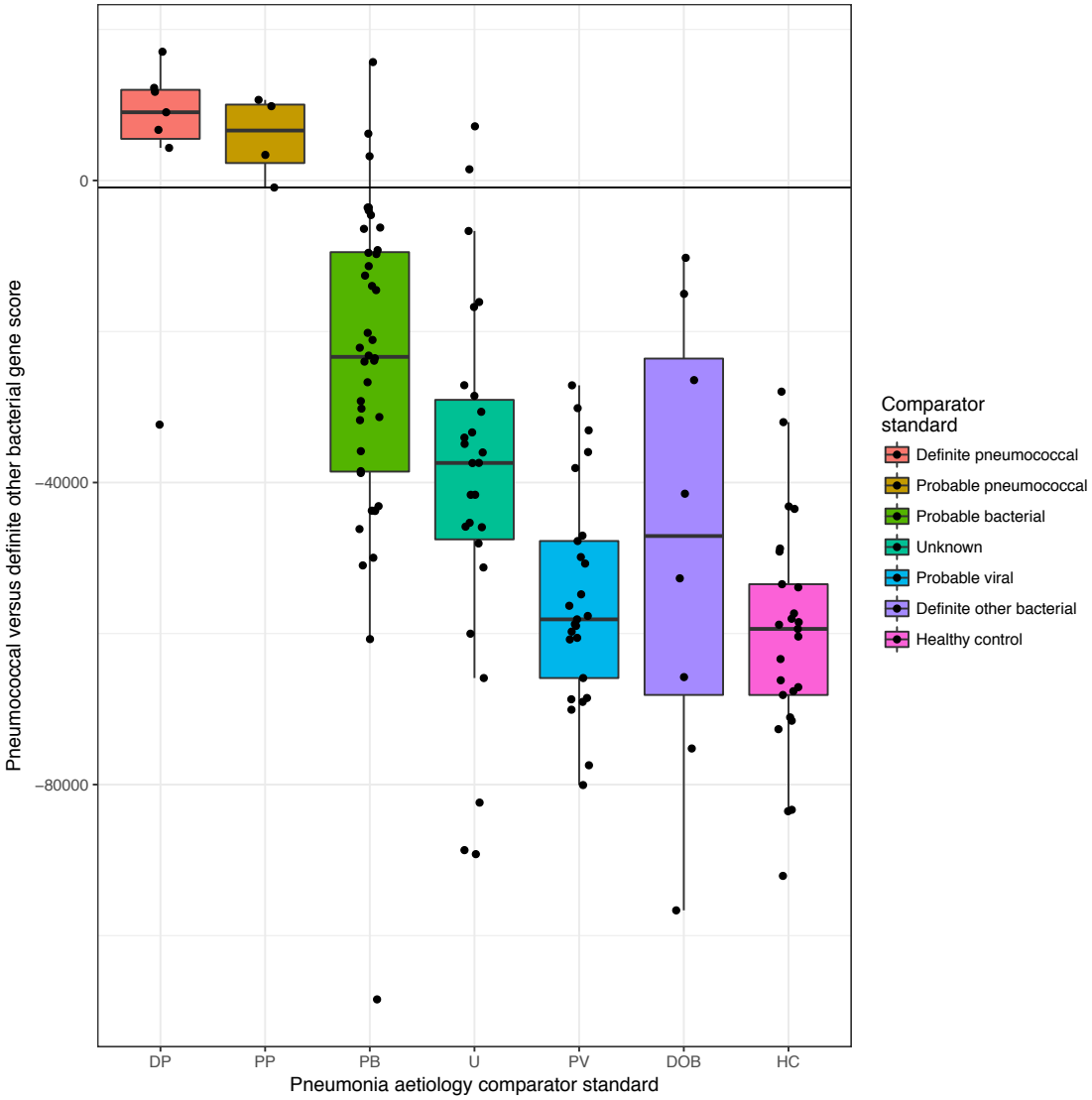


Figure 6.25: Boxplots showing the Pneumococcal Gene Score applied to children with pneumonia of all comparator standards and healthy controls. The optimal threshold (as determined by the Youden Index) was determined from ROC curve analysis of pneumococcal pneumonia vs *definite other bacterial* pneumonia. DP, *definite pneumococcal*; PP, *probable pneumococcal*; PB, *probable bacterial*; U, *unknown*; PV, *probable viral*; DOB, *definite other bacterial*; HC, healthy control.

infection. One might hypothesise that this provides a mechanistic basis for the association between RSV infection in infancy and other diseases, including asthma, in later childhood[281].

6.4.2 Immunological pathways are upregulated in children with pneumonia

In this study, pathways related to immunological activation were upregulated in children with pneumonia vs healthy controls and in acute samples from children with pneumonia vs convalescence. Upregulation of immune-related pathways in response to pneumonia is likely to be an adaptive response to infection and was previously noted by other studies of the host transcriptome response to pneumonia[185, 186, 200]. Nonetheless, the transcriptome in severe infection may be highly labile over time, and in response to differing pathogens, with a hypoimmune response to sepsis associated with poor outcome[171, 172]. This may limit the ability of the transcriptomic response to predict outcome, or diagnose the aetiology of infection, based on a sample taken at a single time point[202].

In this study, the most upregulated pathways in children with pneumonia vs healthy controls were related to neutrophil degranulation, the innate immune system and toll-like receptor signalling. However, when a paired analysis was undertaken – comparing acute and convalescent samples from the same child – the upregulation of interferon α/β signalling became apparent, in addition to neutrophil degranulation, innate immunity and the complement cascade.

Interferon α/β signalling was similarly important when the analysis was limited to the small number of children with pneumococcal pneumonia on whom acute and convalescent samples were available. Neutrophil degranulation, innate immune activation and complement activation was expected given their well-described roles in the immune response against bacterial infections. Upregulation of interferon α/β signalling, presumably derived from neutrophils, was less expected and has typically been associated with the immune response to viral infection. It may be that a proportion of these children had pneumococcal-viral co-infection. An alternative hypothesis is that upregulation of α/β signalling is a maladaptive response to NP carriage of pneumococci, hence their progression to pneumococcal disease. The most parsimonious hypothesis is that a neutrophil-derived interferon α/β signalling pathway is important in protection from

pneumococcal infection. Indeed, this pathway has been described as an important regulator of the global immune response to tuberculosis in adults[190]. The biologic importance of interferon α/β signalling has also been described in a murine model of pneumococcal pneumonia, with mice deficient for interferon α receptor exhibiting enhanced transmigration of pneumococci across the lung into the bloodstream. This effect was reversible with interferon β administration or induction of interferon α with synthetic double-stranded RNA, suggesting novel therapeutic opportunities for the treatment of pneumococcal pneumonia[282].

6.4.3 A distinct transcriptomic signature that discriminates between bacterial and viral pneumonia

I investigated differential gene expression between comparator standards with "known" pneumonia aetiology to develop a transcriptomic signature of bacterial vs viral pneumonia. I developed this transcriptomic signature as a "Bacterial Gene Score", identified an optimal threshold for the score and applied this threshold to the gene score from all acute samples from children with pneumonia and healthy controls.

Based on these data, the optimal Bacterial Gene Score had a sensitivity of 0.68 and specificity of 1.0 for the diagnosis of bacterial pneumonia. Cross validation showed that this score was poor at correctly classifying children with *definite other bacterial* pneumonia, with 6/8 children mis-classified as "viral" pneumonia (Table 6.3). This may reflect the diversity of the transcriptomic response to infection with pathogens such as *S. aureus* (two children), *S. enterica* var Typhi (two children), *Acinetobacter* spp. (two children), *E. coli* or *N. meningitidis* (one child each), which is not correctly accounted for in the Bacterial Gene Score that was otherwise derived from children with pneumococcal pneumonia. Alternatively, mis-classification of the majority of children with *definite other bacterial* pneumonia as "viral" may simply be a reflection of the the lower inflammatory response (as measured by CRP concentration and neutrophil count) in these patients. The lower inflammatory response may be perhaps best described as a shorter molecular distance to health (see above). Indeed, although these children were clinically diagnosed

with pneumonia at presentation to Patan Hospital, they may not be true pneumonia cases – which may be reflected in the transcriptome. Finally, although all of the isolates from children diagnosed with *definite other bacterial* pneumonia were *a priori* defined as pathogens, some may have been contaminants (notably the pathogenesis of *Acinetobacter* spp. is questionable). Nonetheless, the cross validated Bacterial Gene Score classified 50% of children with *probable bacterial* pneumonia, 20% of children with *unknown* aetiology, no children with *probable viral* pneumonia and no healthy controls as "bacterial" (Figure 6.20 and Table 6.3).

Transcriptomic responses have been previously used to discriminate between bacterial and viral infections. In many of these datasets, bioinformatic techniques have been used to remove highly correlated transcripts to yield a small number of genes whose differential regulation maximally discriminates between groups. A smaller, fixed, transcriptomic signature may facilitate the use of alternative platforms, such as quantitative PCR, for the analysis of differential gene expression. A recent study described a two gene transcriptomic (*FAM89A* and *IL144L*) signature that was able to discriminate between bacterial and viral infections in children presenting to hospitals in the European Union and USA with 100% sensitivity and 96% specificity, and additionally validated on children with meningococcal disease (sensitivity 92%) and with non-infectious inflammatory diseases (specificity 96%)[112]. A concurrently published study derived a separate 10 gene signature to discriminate between bacterial and viral infection in febrile infants <60 days of age with >85% sensitivity and specificity – a significant improvement on clinical classification schemes[183]. The two gene transcriptomic signature (derived from children, above) was retrospectively applied to this cohort of febrile infants <60 days of age with a sensitivity of 89% and specificity of 94% to discriminate between bacterial and viral infection.

In summary, I developed a Bacterial Gene Score that may discriminate between bacterial and viral pneumonia, but a larger total sample size, with a larger number of children infected with each bacterial pathogen, and potentially a repeat sample during the acute phase of illness, may better assess its utility. Future work might also use further bioinformatic methods to reduce the number of transcripts used to discriminate between

bacterial and viral pneumonia. Of note, *FAM89A* and *ILI44L*, previously identified as predictors of bacterial and viral infection children[112] were also differentially regulated in these data. The data described here, when examined in the context of previous studies, suggests that the host transcriptome may provide important data to discriminate between bacterial and viral pneumonia, and between healthy carriage of a potential pathogen and disease caused by that pathogen.

6.4.4 A distinct transcriptomic signature that discriminates between pneumococcal and other bacterial pneumonia

I sought to go beyond a comparison of bacterial vs viral pneumonia to assess the ability of a transcriptomic signature to identify children with previously undiagnosed pneumococcal pneumonia. An important early observation, was that children with a high CRP concentration and NP carriage of serotype 1 pneumococci (*probable pneumococcal* pneumonia) had transcriptomes that clustered closely with the transcriptomes of children with *definite pneumococcal* pneumonia.

As noted in the Results section, samples from these children were then aggregated for further analyses. I then derived a Pneumococcal Gene Score from differential gene expression between children with pneumococcal pneumonia and *definite other bacterial* infection, identified an optimal threshold for the score and applied this threshold to the gene score from all acute samples from children with pneumonia and controls. Based on these data, the optimal score had a sensitivity of 0.91 and a specificity of 1.0 for pneumococcal pneumonia, which was robust to cross validation, and identified an additional 3/40 children with *probable bacterial* pneumonia and and 2/26 children with *unknown* aetiology as pneumococcal pneumonia cases.

The derivation of a transcriptomic signature for pneumococcal pneumonia by a comparison with *definite other bacterial* pneumonia is highly stringent, since one might hypothesise that the transcriptome of a child with pneumococcal pneumonia may be more similar to that of a child with *definite other bacterial* pneumonia than, for example, viral pneumonia. Of note, the microbiological methods rigorously assessed in the PERCH

study had a lower sensitivity and specificity for the discrimination of pneumococcal pneumonia from NP carriage of pneumococci in healthy children (see Chapter 1, Table 1.2), and were unable to discriminate between children with pneumonia. One interesting approach will be to integrate the transcriptomic results described here with proxy markers of pneumococcal pneumonia in a Bayesian manner to predict pneumonia aetiology in individual children.

The Pneumococcal Gene Score developed here shows proof of principle that transcriptomic analyses can discriminate between children with pneumococcal pneumonia and other aetiologies of pneumonia. This supports data from previous studies to identify pathogen-specific transcriptomic signatures in childhood respiratory infections such as RSV[198, 200], and tuberculosis[190, 191]. A further application of the use of a transcriptomic signature for pneumococcal pneumonia will be to assess the impact of pneumococcal conjugate vaccination on pneumonia aetiology in children in Nepal. As described in Chapter 1, assessing the impact of 10-valent PCV is of public health importance in Nepal.

A further issue is the choice of method used to calculate the optimal threshold for a diagnostic score: for example, acceptance of a small loss of specificity would classify a considerably larger proportion of children with *probable bacterial* pneumonia as "pneumococcal" on the basis of the Pneumococcal Gene Score. Potential future work may use bioinformatic methods to limit the number of genes used to classify pneumococcal pneumonia. In a larger set of samples, a set of genes derived from a comparison between acute pneumococcal pneumonia and convalescence (Table 6.5) may provide a rigorous transcriptomic signature of pneumococcal pneumonia.

6.4.5 Limitations and summary

As noted below, this analysis is limited by sample size – and in particular the number of children with pneumonia of "known" aetiology. Additionally, the diversity of transcriptomic responses of children with *definite other bacterial* pneumonia suggests that increased numbers of children with non-pneumococcal bacterial infections, but of similar severity to

pneumococcal pneumonia would provide increased power for the derivation of the Bacterial Gene Score, and provide further testing of the Pneumococcal Gene Score.

In previously published studies, the age of a child does not appear to affect the ability of transcriptomic signatures to discriminate between bacterial and viral infection[112, 183, 184]. I was unable to formally adjust for age in these multivariable regression models, since the data were limited by sample size, and limited overlap between the ages of children with pneumococcal pneumonia and *definite other bacterial* pneumonia (see also Chapter 3).

The principal limitation of any study to assess pneumonia aetiology is the lack of a "gold standard" diagnostic test against which to assess a novel diagnostic test. I attempted to offset this difficulty with the development of comparator standards of pneumonia aetiology. As noted in previous chapters, the criteria for the development of these comparator standards were based on predictors for invasive pneumococcal disease in children in the Inpatient Pneumonia Study in classification tree analyses and published literature on thresholds for bacterial infection in children[111, 112]. These comparator standards are used consistently across this thesis – but other comparator standards may have validity, such as comparisons of children with and without endpoint consolidation on chest radiograph.

Here, I have developed transcriptomic signatures of bacterial and pneumococcal pneumonia. Key further analyses will assess whether these tests provide additive value to currently used diagnostic tests for the aetiology of pneumonia. The data presented here suggest that transcriptomic analyses may be usefully integrated with current state-of-the-art methods – possibly through a Bayesian modelling approach[74] – to better predict the aetiology of childhood pneumonia.

For, in the final analysis, our most basic common link is that we all inhabit this small planet. We all breathe the same air. We all cherish our children's future. And we are all mortal.

— John F Kennedy, Speech at The American University, Washington DC, 10th June 1963

7

Summary: childhood pneumococcal pneumonia at Patan Hospital, Nepal

Pneumonia is the greatest cause of childhood mortality outside the neonatal period, and yet the pathogen-specific aetiology of disease remains poorly defined[1, 10]. The lack of accurate diagnostic tests for pneumonia aetiology leads to unnecessary antibiotic administration to children with pneumonia of viral aetiology, contributing to the rise in antimicrobial resistance[283], and inhibits the rational prioritisation of pathogen-specific vaccine development and implementation[12]. Considerable progress has been made in reducing childhood mortality from pneumonia, with the global number of deaths in children <5 years of age, nearly halving from 1.7 million in the late 1990s to ∼920 000 by 2015[1, 8, 9]. In part, this is attributable to the implementation of vaccination against *H. influenzae* type b, and more recently against *S. pneumoniae*, particularly in low and middle income countries[10, 11].

Analyses based on data from vaccine probe studies suggest that pneumococci are

likely to remain a significant cause of childhood pneumonia – and associated mortality – in settings that have not introduced pneumococcal conjugate vaccines (PCV)[11]. In January 2014, these settings included all South Asian countries with the exception of Afghanistan and Pakistan[12]. Nepal introduced 10-valent PCV during 2015, with the implementation into the immunisation schedule of infants in the Kathmandu Valley in August 2015.

Working within a team at Patan Hospital, Kathmandu, Nepal, I used data from a long-term childhood IBD surveillance programme to describe the prevalence of invasive pneumococcal disease in children with suspected IBD admitted to Patan Hospital between March 2005 and December 2016. In the absence of sensitive tests for pneumococcal infection in children with pneumonia, proxy markers of pneumococcal pneumonia were developed by examining associations between clinical data from children with pneumonia admitted to Patan Hospital since March 2014 (the Inpatient Pneumonia Study) and invasive pneumococcal disease. This allowed me to assess the early impact of 10-valent PCV on proxy markers of pneumococcal pneumonia, and on nasopharyngeal (NP) carriage of pneumococcal serotypes, in these children.

Current state-of-the-art diagnostic methods for pneumonia aetiology seek to isolate and culture the causative organism, or identify pathogen antigens or nucleic acids. These tests are either insensitive for pathogens (e.g. culture of blood)[74], or cannot distinguish healthy NP carriage from disease for most pathogens[46, 47, 82–85]. I hypothesised that examination of the immune response in children with pneumonia may be used to discriminate between differing aetiologies of pneumonia with increased accuracy than tests based on direct identification of pathogens. I investigated two recently developed approaches for the aetiological diagnosis of infection: assay of antibodies from lymphocyte supernatant (ALS), and analysis of differential gene expression to develop transcriptomic signatures of disease.

Here, I summarise the results of these investigations, and briefly highlight potential implications of these results for future work on childhood pneumonia in Nepal and elsewhere.

7.1 Pneumococci are the most common invasive pathogens in childhood pneumonia

Pneumococci were the second-most prevalent pathogens isolated from the blood of children admitted to Patan Hospital with suspected IBD, after *S. enterica* var Typhi, and were the most prevalent pathogen isolated from children with a clinical diagnosis of pneumonia between March 2005 and December 2016. Between 2010 and 2016, where detailed data on all admissions was available, the prevalence of invasive pneumococcal disease was 0.6% of all children enrolled (17% of positive blood cultures). Similar results have been described in other low income South and South East Asian settings prior to the introduction of PCV[77, 234, 235]; higher prevalence of childhood invasive pneumococcal disease has been described in African settings associated with a higher prevalence of HIV infection[235, 236].

There was an apparent decline in the prevalence of invasive pneumococcal disease from 2009 onwards. This decline was predominantly in infants, such that since 2014 the majority of invasive pneumococcal disease has been identified in children $\geq$5 years of age – an age group in whom invasive pneumococci are typically of serotypes contained within 10-valent PCV (and in particular, serotype 1). The relatively low prevalence of invasive pneumococcal disease, in contrast to the hypothesised burden of pneumococcal pneumonia on the basis of vaccine probe studies, and the age distribution of invasive pneumococcal disease at Patan Hospital, make this an insensitive endpoint for the assessment of the impact of 10-valent PCV introduction to the infant immunisation schedule in Nepal. I therefore developed proxy markers for pneumococcal pneumonia by examining detailed data from children admitted to Patan Hospital with a clinical diagnosis of pneumonia (all of whom were also enrolled in the childhood IBD surveillance programme). Radiographic endpoint consolidation (OR 19), increased plasma CRP concentration (OR 1.1 per 10 mg/dl increase), and NP carriage of serotype 1 pneumococci (OR 5.3) were all associated with invasive pneumococcal disease when adjusted for age.

A classification tree analysis was used to develop a classification scheme for the probability of invasive pneumococcal disease in children admitted to Patan Hospital with a clinical diagnosis of pneumonia. The optimal model used plasma CRP concentration, NP carriage of serotype 1 pneumococci, and blood neutrophil count. This was modified (in particular with the optimisation of the threshold for CRP concentration) to develop comparator standards of pneumonia aetiology, used extensively to assess novel diagnostic tests for pneumococcal pneumonia (discussed below).

7.2 Early impact of 10-valent PCV in children with pneumonia

10-valent PCV was introduced to the infant immunisation schedule in Kathmandu Valley in August 2015 at 6 weeks, 10 weeks and 9 months of age. There was a limited "catch-up" campaign in children $\geq$10 weeks and <1 year of age.

Immediately prior to introduction of 10-valent PCV, in late April 2015, a series of large earthquakes caused severe damage to Kathmandu and central Nepal, with the direct loss of 8000 lives and displacement of approximately 2.5 million people from their homes. Large fluctuations of people, and the formation of internally-displaced people's camps, followed. In addition, difficulties importing clean petroleum fuel from India led to a significant increase in ambient air pollution in the Kathmandu Valley during the winter of 2015/2016. Both of these events are likely to have increased transmission of respiratory pathogens and presentation to hospital with pneumonia in children[16, 212].

These two extreme events were unable to be adjusted for, or modelled as separate "interruptions" in interrupted time series analyses, since they were immediately adjacent to introduction of 10-valent PCV, nor was seasonal variation in pneumonia admissions due to the short pre-introduction period. Analyses using clinical phenotypes of pneumonia, and proxy markers of pneumococcal pneumonia, showed a 4% increase in the odds of admission of young children (<5 years of age) with fast breathing pneumonia, a 3% increase in chest indrawing pneumonia and a 4% increase in pneumonia with danger

signs (all according to WHO definitions[15]) vs the odds of admission to Patan Hospital with a non-pneumonia diagnosis. There was no change in the odds of admission with pneumonia with radiographic endpoint consolidation. In addition there was no change in odds of NP carriage of vaccine-type pneumococci over NP carriage of non-vaccine-type pneumococci.

Given the difficulties isolating the impact of 10-valent PCV introduction from these extreme events in Nepal, I used a separate method – the indirect cohort method – to directly assess vaccine effectiveness against the same clinical phenotypes of pneumonia and proxy markers of pneumococcal pneumonia. The indirect cohort method, based on actual receipt of vaccination rather than trends over time, estimated vaccine effectiveness to be 84% against NP carriage of vaccine-type pneumococci. There was no significant effect on admission with clinical phenotypes of pneumonia or pneumonia with endpoint consolidation.

The accuracy of both interrupted time series analyses and the indirect cohort method to assess the impact of 10-valent PCV on childhood pneumonia would be improved by the enrolment of further children – and work is currently being done to effect this. In particular, interrupted time series analyses require a longer post-introduction period (and ideally, pre-introduction period) in the context of seasonal variations in pneumonia admissions and two extreme events temporally associated with the introduction of 10-valent PCV.

7.3 IgG ALS to pneumococcal proteins for aetiological diagnosis

A key difficulty to assessing the impact of 10-valent PCV on pneumococcal pneumonia in Nepal is the lack of an accurate diagnostic test for the aetiology of childhood pneumonia[2]. An accurate diagnostic test for pneumonia aetiology would also be of considerable clinical use, potentially limiting unnecessary administration of antibiotics, and informing the rational development of vaccines against pathogens implicated in childhood pneumonia.

I investigated whether assay of ALS had utility for the aetiological diagnosis of pneumonia in children at Patan Hospital. This technique uses the supernatant from cultured suspensions of peripheral blood mononuclear cells (without extrinsic stimulation) isolated from the blood of patients with disease, a proportion of which are likely to be antigen-specific antibody-secreting cells (ASCs) during acute illness[121, 138]. Antibody concentrations to specific antigens in the resulting supernatant (ALS) can then be assayed. ALS was generated from the blood of children with pneumonia and IgG ALS to 13 pneumococcal capsular polysaccharides (PS) and 5 pneumococcal proteins was assayed in fluorescence-based multiplexed immunoassays (FMIA).

Assay of IgG ALS to pneumococcal capsular polysaccharides revealed an apparent non-specific ASC response against most pneumococcal capsular polysaccharides (particularly PS3), and the capsular polysaccharide of *H. influenzae* type b, that was unable to discriminate between children with *definite pneumococcal* pneumonia, and other comparator standards. However, detailed laboratory experiments were unable to delineate the exact cogent antigen for these antibodies.

In contrast assay of ALS to pneumococcal proteins, particularly choline binding protein A (CbpA) and pneumococcal histidine triad D (PhtD), had some utility in discriminating between children with *definite pneumococcal* pneumonia and other comparator standards. In analyses of all 348 children enrolled to the study on whom ALS samples were available, assay of IgG ALS to PhtD discriminated between children with *definite pneumococcal* and *definite other bacterial* pneumonia with a sensitivity of 88% and specificity of 71%. Infants and young children may produce a limited adaptive immune response. I therefore excluded children <2 years of age from analyses: in this very limited data set CbpA was the best-forming antigen, with assay of IgG ALS able to discriminate between *definite pneumococcal* and *definite other bacterial* pneumonia with a sensitivity of 87.5% and specificity of 67%.

The sensitivity of assay of ALS is likely to be limited by the transient nature of circulation of antigen-specific ASCs in peripheral blood in children with infections[121]. Repeated sampling time points during acute illness may increase sensitivity of these methods, as

may the assay of ALS to a broader array of pneumococcal (and potentially other pathogen) antigens. Increasing the specificity of assay of ALS may be more difficult, since ASCs that circulate during acute illness are hypothesised to have been recently activated and therefore express non-affinity matured B cell receptors and antibodies.

7.4 Transcriptomics for aetiological diagnosis

Differential gene expression analyses have been used to discriminate between patients with a number of infectious disease aetiologies[189–191, 198], and between bacterial and viral infections in children with high accuracy[112, 183–185, 277]. I hypothesised that these techniques may therefore be able to discriminate between bacterial and viral pneumonia in children at Patan Hospital, and potentially discriminate between children with pneumococcal and other bacterial pneumonia. If so, this could provide proof-of-concept for the development of a novel diagnostic test, and be used to assess the impact of 10-valent PCV on pneumococcal pneumonia in children in Nepal. RNA sequencing was therefore performed on acute samples from 110 children at Patan Hospital with pneumonia, and convalescent samples from 57 of these children, to generate gene expression data.

Differential gene expression was used to develop a transcriptomic signature of bacterial pneumonia (the Bacterial Gene Score), which discriminated between children with bacterial pneumonia and viral pneumonia with a sensitivity of 68% and 100% specificity at an optimal threshold. However, cross validation showed that this signature performed poorly at differentiating children with *definite other bacterial* pneumonia from children with viral pneumonia – possibly due to the small number of children infected with diverse pathogens in these comparator standards.

Using similar methods, a transcriptomic signature (the Pneumococcal Gene Score) was able to discriminate between children with pneumococcal pneumonia and *definite other bacterial* pneumonia with a sensitivity of 91% and specificity of 100% at optimal an threshold, which was supported by cross validation. Application of this threshold across

comparator standards revealed 8% of children with *probable bacterial* pneumonia and 8% of children with *unknown* aetiology to have a "positive" pneumococcal gene score. However, this proportion would increase substantially with small changes to the threshold for a positive pneumococcal gene score.

Differential gene expression analyses between acute and convalescent samples from children with pneumonia (all cause) revealed a neutrophil-derived interferon α/β signalling pathway that was particularly pronounced in children with pneumococcal pneumonia – although cases of *definite other bacterial* were to few to formally assess its pathogen-specificity. This is in concordance with experimental murine models of pneumococcal pneumonia and may represent an adaptive host immune response to limit pneumococcal transmigration across lung epithelia[282].

Previous work has not formally described a diagnostic transcriptomic signature for pneumococcal infection in children with pneumonia. However, these results are in concordance with the use of transcriptomic analysis to diagnose childhood tuberculosis[191] and to discriminate between bacterial and viral infection in children[112, 183–185, 277]. In these analyses I did not use bioinformatic methods to remove highly correlated transcripts (and thereby produce a limited set of discriminatory genes).

The use of transcriptomic techniques therefore has considerable potential for the aetiological diagnosis of pneumonia. Further work on developing, and trialling, the use of limited gene sets may yield a more practical platform on which to assess transcriptomic signatures of to discriminate between pneumonia aetiologies in a clinically-relevant timescale.

7.5 Implications for future studies of pneumonia aetiology

This thesis describes data on the prevalence of invasive pneumococcal disease and proxy markers of pneumococcal pneumonia in children admitted to Patan Hospital with pneumonia, assesses the early impact of 10-valent PCV on proxy markers of pneumococcal pneumonia and investigates new diagnostic tests for pneumococcal

pneumonia. These data are all from a single site in Nepal. Nonetheless, the methods used within this work could be clearly adapted to the identification of other pathogens, to other sites, and indeed to other infectious disease syndromes in children or adults.

The PERCH study is the most comprehensive study of childhood pneumonia ever undertaken, and has rigorously assessed the utility of diagnostic tests to identify the pathogen-specific aetiology of pneumonia in individual children at the epidemiological level[10]. To integrate the results of numerous diagnostic tests, with varying sensitivity and specificity for pathogens, a partial latent class model, using a Bayesian statistical approach, has been formalised[74]. However, the limited ability of diagnostic tests used with the PERCH study to discriminate between pneumonia aetiologies will limit the interpretability of results. Future studies of pneumonia aetiology, or of vaccine impact more generally, may benefit from the use of diagnostic tests based on the host immune response (in particular, transcriptomic analyses), in addition to current state-of-the-art microbiological methods, for the accurate aetiological diagnosis of infectious disease. The use of the statistical approach developed for the PERCH study to integrate these diagnostic test results is likely to be of considerable importance in future.

The difficulty in ascribing an aetiological diagnosis to cases of childhood pneumonia is a hindrance to the prioritisation of healthcare interventions, at an individual and public health level. The analysis of the early data on the impact of 10-valent PCV on childhood pneumonia at Patan Hospital exemplifies this. In the absence of detailed demographic data at a population level, analysis of hospitalisation data was required to assess the impact of 10-valent PCV. Improved diagnostic tests, such as the use of transcriptomic analysis, may aid these impact analyses in future. More specifically, given the short post-vaccine period, age distribution of childhood invasive pneumococcal disease and age distribution of proxy markers of pneumococcal pneumonia, a longer post-vaccine period of observation is undoubtedly required.

Appendices

A

Appendix

A.1 General methods

Figure A.1 details the methods used for the generation of ALS in the PANDIT Study.

A.2 Transcriptomic investigation of childhood pneumonia

A.2.1 Bioinformatic methods: scripts

Please contact Michael Carter (michael.carter@paediatrics.ox.ac.uk) or Wanseon Lee (wl@well.ox.ac.uk) for access to scripts.

A.2.2 Mapping of reads

https://github.com/jknightlab/ngs_guidelines/blob/master/RNASeq/2.tophat2_mapping.sh

Oxford Vaccine Group Laboratory (OVGL)

PANDIT

Pneumonia Aetiology in Nepal: Diagnosis with Immunology and Transcriptomics

QUICK NOTES PBMC and ALS Preparation and Harvest (version 3.0)
Gentle handling during this preparation ensures healthy cells. Vigorous pipetting should be avoided. Do all procedures in the laminar flow cabinet.

PBMC Separation
1 – Add 5ml of Histopaque 1077 to a 15ml centrifuge tube.

2 – Layer the blood onto 5ml of Histopaque 1077 with a pipetter on slow speed or Pasteur pitpette.

3 – Centrifuge (400 g, 20 mins with the low acceleration and the brake off).

4 – Following centrifugation, remove the plasma down to just above the PBMC layer using a sterile Pasteur pipette, then store the plasma (labelled cryovials at -80°C).

5 – Using a fresh sterile Pasteur pipette remove the PBMCs to a fresh 15ml tube and top up with R0 to 2ml. Remove 50µl of cells for counting.

6 – Remove Histopaque from Falcon tube, and remove red cell pellet, label and store at –80ºC.

7 – a) Wash 1: Top up PBMC aliquot volume to 5ml with R0 and spin the sample 2000 RPM for 10 mins
b) Carefully tip away and then suction remaining R0 from the cell pellet
c) Add a further 5 ml of R0 and resuspend the cell pellet fully with gentle suctioning with a Pasteur pipette, and mix gently but thoroughly
d) Remove 50 µl of cell suspension for cell counting (see step 8)
e) Spin the sample at 2000 RPM for a further 10 mins (whilst doing step 8)

8 – To count cells take 50ul of cells + 50ul PBS + 50ul trypan blue, mix and transfer ~ 10µl to the plastic hemocytometer.

Example : Count (in 16 small squares) 1/2/3 = 45/50/52 = 49 x 3 (dilution factor) = 1.47 x 10^6/ml

To calculate the total number of cells multiply by total volume at time of cell count (i.e. 5ml)
= 1.47 x 10^6/ml x 5 = 7.35 x 10^6 total

9 – Remove all R0 from the cell pellet by tipping the tube (not suctioning). Resuspend the cell pellet in R10 to a density of 1 x10^7/ml as below.

To calculate the volume of R10 required divide your total cell concentration by 10
= 7.35 ÷ 10 = 0.735 ml (so you would resuspend your cell pellet in 0.735 ml of R10)

10 – Transfer your cell suspension to a single well of a 24 well plate (add sterile PBS to the remaining empty wells to avoid evaporation during culture). If the cell suspension is ≤300 µL volume, then use the well of a 96 well plate.

11 – Incubate the plate for 48 hours at 37°C with CO_2 sachets (complete ALS sample sheet).

Form Title: Quick notes: ALS sample preparation and harvest for PANDIT
v.3.0 14th January 2016, Michael Carter

Page 1 of 2

Figure A.1: Methods for the generation of ALS in the pneumonia studies at Patan Hospital.

A.2.3 Filtering of reads

https://github.com/jknightlab/ngs_guidelines/blob/master/RNASeq/3.2.filtering.genome.sh

A.2.4 Generation of gene counts and edgeR pipeline

https://github.com/jknightlab/ngs_guidelines/blob/master/RNASeq/PANDIT_RNASeq_edgeR_pipeline

Bibliography

[1] Liu, L., Oza, S., Hogan, D., et al. “Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals”. *Lancet* 388.10063 (2016), pp. 3027–3035.

[2] Murdoch, D. R., O’Brien, K. L., Driscoll, A. J., et al. “Laboratory methods for determining pneumonia etiology in children”. *Clin Infect Dis* 54 Suppl 2 (2012), S146–52.

[3] O’Brien, K. L., Wolfson, L. J., Watt, J. P., et al. “Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates”. *Lancet* 374 (2009), pp. 893–902.

[4] Kelly, D. F., Thorson, S., Maskey, M., et al. “The burden of vaccine-preventable invasive bacterial infections and pneumonia in children admitted to hospital in urban Nepal”. *Int J Infect Dis* 15.1 (2011), e17–23.

[5] Williams, E. J., Thorson, S., Maskey, M., et al. “Hospital-based surveillance of invasive pneumococcal disease among young children in urban Nepal”. *Clin Infect Dis* 48 Suppl 2 (2009), S114–22.

[6] Chhetri, U. D., Shrestha, S., Pradhan, R., et al. “Clinical profile of invasive pneumococcal diseases in Patan Hospital, Nepal”. *Kathmandu Univ Med J* 9.33 (2011), pp. 45–9.

[7] Collaborators, GBD2015 Child Mortality. “Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015”. *Lancet* 388 (2016), pp. 1725–74.

[8] You, D., Hug, L., Ejdemyr, S., et al. “Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation”. *Lancet* (2015).

[9] Lozano, R., Naghavi, M., Foreman, K., et al. “Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010”. *Lancet* 380 (2012), pp. 2095–2128.

[10] O’Brien, K. L., Baggett, H. C., Brooks, W. A., et al. “Introduction to the Epidemiologic Considerations, Analytic Methods, and Foundational Results From the Pneumonia Etiology Research for Child Health Study”. *Clin Infect Dis* 64.S3 (2017), S179–S184.

[11] Wahl, B., O’Brien, K. L., Greenbaum, A., et al. “Global, regional, and national burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b in children in the era of conjugate vaccines: updated estimates from 2000–2015” ().

[12] IVAC. *VIMS Report: Global Vaccine Introduction (January 2015)*. Baltimore, USA: Johns Hopkins Bloomberg School of Public Health, 2015.

[13] Nunes, M. C., Gottberg, A. V., Gouveia, L. de, et al. “The impact of antiretroviral treatment on the burden of invasive pneumococcal disease in South African children: a time series analysis”. *AIDS* 25 (2011), pp. 453–62.

[14] Roux, D. M. le, Myer, L., Nicol, M. P., et al. “Incidence and severity of childhood pneumonia in the first year of life in a South African birth cohort: the Drakenstein Child Health Study”. *Lancet Glob Health* 3 (2015), e95–103.

[15] WHO. *Revised WHO classification and treatment of childhood pneumonia at health facilities: evidence summaries*. Geneva, Switzerland: WHO, 2014.

[16] Cohen, A. J., Brauer, M., Burnett, R., et al. “Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015”. *Lancet* 389.10082 (2017), pp. 1907–1918.

[17] Adegbola, R. A. and Levine, O. S. “Rationale and expectations of the Pneumonia Etiology Research for Child Health (PERCH) study”. *Expert review of respiratory medicine* 5.6 (2011), pp. 731–733.

[18] WHO. *Improving child health IMCI: the integrated approach*. Geneva, Switzerland: WHO, 1999.

[19] Addo-Yobo, E., Chisaka, N., Hassan, M., et al. “Oral amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59 months: a randomised multicentre equivalency study”. *Lancet* 364 (2004), pp. 1141–8.

[20] Hazir, T., Fox, L. M., Nisar, Y. B., et al. “Ambulatory short-course high-dose oral amoxicillin for treatment of severe pneumonia in children: a randomised equivalency trial”. *Lancet* 371 (2008), pp. 49–56.

[21] Subhi, R., Adamson, M, Campbell, H., et al. “The prevalence of hypoxaemia among ill children in developing countries: a systematic review”. *Lancet Infect Dis* 9 (2009), pp. 219–29.

[22] Tosif, S. and Duke, T. “Evidence to support oxygen guidelines for children with emergency signs in developing countries: a systematic review and physiological and mechanistic analysis”. *J Trop Pediatr* (2017).

[23] McCollum, E. D., Nambiar, B., Deula, R., et al. “Impact of the 13-Valent Pneumococcal Conjugate Vaccine on Clinical and Hypoxemic Childhood Pneumonia over Three Years in Central Malawi: An Observational Study”. *PLoS One* 12.1 (2017), e0168209.

[24] Fleming, S., Thompson, M., Stevens, R., et al. “Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies”. *Lancet* 377.9770 (2011), pp. 1011–1018.

[25] Anon. *The Pneumonia Etiology Research for Child Health (PERCH) Protocol v.2.0*. 2012.

[26] Crawley, J., Prosperi, C., Baggett, H. C., et al. “Standardization of Clinical Assessment and Sample Collection Across All PERCH Study Sites”. *Clin Infect Dis* 64.S3 (2017), S228–S237.

[27] Selwyn, BJ. “The epidemiology of acute respiratory tract infection in young children: comparison of findings from several developing countries”. *Rev Inf Dis* 12.Supplement 8 (1990), S870–S888.

[28] Moyes, Jocelyn, Cohen, Cheryl, Pretorius, Marthi, et al. “Epidemiology of respiratory syncytial virus–associated acute lower respiratory tract infection hospitalizations among HIV-infected and HIV-uninfected South African children, 2010–2011”. *Journal Infect Dis* 208.suppl 3 (2013), S217–S226.

[29] Chheng, K., Carter, M. J., Emary, K., et al. “A prospective study of the causes of febrile illness requiring hospitalization in children in Cambodia”. *PLoS One* 8.4 (2013), e60634.

[30] Cherian, T., Mulholland, E. K., Carlin, J. B., et al. “Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies”. *Bull WHO* 83.5 (2005), pp. 353–359.

[31] Hansen, J., Black, S., Shinefield, H., et al. “Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than 5 years of age for prevention of pneumonia: updated analysis using World Health Organization standardized interpretation of chest radiographs”. *Pediatr Infect Dis J* 25.9 (2006), pp. 779–81.

[32] Klugman, K. P., Madhi, S. A., Huebner, R. E., et al. “A trial of a 9-valent conjugate vaccine in children with and those without HIV infection”. *N Engl J Med* 349.14 (2003), pp. 1341–8.

[33] Cutts, F. T., Zaman, S. M. A., 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* 365 (2005), pp. 1139–1146.

[34] Lucero, M. G., Nohynek, H., Williams, G., et al. “Efficacy of an 11-valent pneumococcal conjugate vaccine against radiologically confirmed pneumonia among children less than 2 years of age in the Philippines: a randomized, double-blind, placebo-controlled trial”. *Pediatr Infect Dis J* 28.6 (2009), pp. 455–62.

[35] Feikin, D. R., Hammitt, L. L., Murdoch, D. R., et al. “The Enduring Challenge of Determining Pneumonia Etiology in Children: Considerations for Future Research Priorities”. *Clin Infect Dis* 64.S3 (2017), S188–S196.

[36] WHO. *Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illnesses*. Geneva, Switzerland: WHO, 2013.

[37] WHO. *Vaccine preventable diseases surveillance: global invasive bacterial surveillance and information bulletin*. Report. WHO, 2015.

[38] Singer, M., Deutschman, C. S., Seymour, C. W., et al. “The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)”. *JAMA* 315.8 (2016), pp. 801–10.

[39] Pasteur, L. “Sur une maladie nouvelle provoquée par la salive d’un enfant mort de rage”. *Comptes Rendus de l’Académie des Sciences de Paris* 92 (1881), p. 159.

[40] Disease Control, Centers for and Prevention. *Laboratory Methods for the Diagnosis of Meningitis Caused by Neisseria meningitidis, Streptococcus pneumoniae, and Haemophilus influenzae*. Atlanta, GA, USA: CDC, 2011.

[41] Institut, Statens Serum. *Streptococcus pneumoniae: textbook in serotyping, virulence factors and enzyme-linked immunosorbent assay (ELISA) for measuring pneumococcal antibodies*. Copenhagen, Denmark: SSI, 2013.

[42] Bentley, S. D., Aanensen, D. M., Mavroidi, A., et al. "Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes". *PLoS Genetics* 2 (2006), e31.

[43] Dillard, J. P., Vandersea, M. W., and Yother, J. "Characterisation of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*." *J Exp Med* 181 (1995), pp. 973–983.

[44] 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* 49.5 (2011), pp. 1784–9.

[45] Moore, C. E., Sengduangphachanh, A., Thaojaikong, T., et al. "Enhanced determination of *Streptococcus pneumoniae* serotypes associated with invasive disease in Laos by using a real-time polymerase chain reaction serotyping assay with cerebrospinal fluid." *Am J Trop Med Hyg* 83 (2010), pp. 451–7.

[46] Morpeth, S. C., Deloria Knoll, M., Scott, J. A. G., et al. "Detection of Pneumococcal DNA in Blood by Polymerase Chain Reaction for Diagnosing Pneumococcal Pneumonia in Young Children From Low- and Middle-Income Countries". *Clin Infect Dis* 64.S3 (2017), S347–S356.

[47] Deloria Knoll, M., Fu, W., Shi, Q., et al. "Bayesian Estimation of Pneumonia Etiology: Epidemiologic Considerations and Applications to the Pneumonia Etiology Research for Child Health Study". *Clin Infect Dis* 64.S3 (2017), S213–S227.

[48] Satzke, C., Dunne, E. M., Porter, B. D., et al. "The PneuCarriage Project: A Multi-Centre Comparative Study to Identify the Best Serotyping Methods for Examining Pneumococcal Carriage in Vaccine Evaluation Studies". *PLoS Med* 12.11 (2015), e1001903; discussion e1001903.

[49] Kandasamy, R., Gurung, M., Thapa, A., et al. "Multi-serotype pneumococcal nasopharyngeal carriage prevalence in vaccine naive nepalese children, assessed using molecular serotyping". *PLoS One* 10.2 (2015), e0114286.

[50] Bogaert, D., Groot, R. de, and Hermans, P. W. M. "*Streptococcus pneumoniae* colonisation: the key to pneumococcal disease". *Lancet Infect Dis* 4.3 (2004), pp. 144–154.

[51] Coles, C. L., Sherchand, J. B., Khatry, S. K., et al. "Nasopharyngeal carriage of *S. pneumoniae* among young children in rural Nepal". *Trop Med Int Health* 14.9 (2009), pp. 1025–33.

[52] Hanieh, S., Hamaluba, M., Kelly, D. F., et al. "*Streptococcus pneumoniae* carriage prevalence in Nepal: evaluation of a method for delayed transport of samples from remote regions and implications for vaccine implementation". *PLoS One* 9.6 (2014), e98739.

[53] Turner, G. D., Bunthi, C., Wonodi, C. B., et al. "The role of postmortem studies in pneumonia etiology research". *Clin Infect Dis* 54 Suppl 2 (2012), S165–71.

[54] Coles, C. L., Kanungo, R., Rahmathullah, L., et al. "Pneumococcal nasopharyngeal colonization in young South Indian infants". *Pediatr Infect Dis J* 20.3 (2001), pp. 289–95.

[55] Devi, U., Ayyagari, A., Devi, K. R., et al. “Serotype distribution & sensitivity pattern of nasopharyngeal colonizing *Streptococcus pneumoniae* among rural children of eastern India”. *Indian J Med Res* 136 (2012), pp. 495–498.

[56] Adegbola, R. A., DeAntonio, R., Hill, P. C., 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* 9.8 (2014), e103293.

[57] Chewapreecha, C., Harris, S. R., Croucher, N. J., et al. “Dense genomic sampling identifies highways of pneumococcal recombination”. *Nat Genet* 46.3 (2014), pp. 305–9.

[58] Golubchik, T., Brueggemann, A. B., Street, T., et al. “Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event”. *Nat Genet* 44.3 (2012), pp. 352–5.

[59] Croucher, N. J., Chewapreecha, C., Hanage, W. P., et al. “Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape”. *Genome Biol Evol* 6.7 (2014), pp. 1589–602.

[60] Iwajomo, O. H., Finn, A., Moons, P., et al. “Deteriorating pneumococcal-specific B-cell memory in minimally symptomatic African children with HIV infection”. *J Infect Dis* 204.4 (2011), pp. 534–43.

[61] Bliss, S. J., O’Brien, K. L., Janoff, E. N., et al. “The evidence for using conjugate vaccines to protect HIV-infected children against pneumococcal disease”. *Lancet Infect Dis* 8 (2007), pp. 67–80.

[62] Iwajomo, O. H., Finn, A., Ogunniyi, A. D., et al. “Impairment of pneumococcal antigen specific isotype-switched IgG memory B-cell immunity in HIV infected Malawian adults”. *PLoS One* 8.11 (2013), e78592.

[63] Bruton, O. C. “Agammaglobulinaemia”. *Pediatrics* 9 (1952), pp. 722–8.

[64] Casanova, J. L. and Abel, L. “Primary immunodeficiencies: a field in its infancy”. *Science* 317 (2007), pp. 617–619.

[65] Scott, J. A., Hall, A. J., Dagan, R., et al. “Serogroup-specific epidemiology of *Streptococcus pneumoniae*: associations with age, sex, and geography in 7,000 episodes of invasive disease”. *Clin Infect Dis* 22 (1996), pp. 973–81.

[66] Brueggemann, A. B., Griffiths, D. T., Meats, E., et al. “Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential”. *J Infect Dis* 187 (2003), pp. 1424–1432.

[67] Greenberg, D., Givon-Lavi, N., Newman, N., et al. “Nasopharyngeal carriage of individual *Streptococcus pneumoniae* serotypes during pediatric pneumonia as a means to estimate serotype disease potential”. *Pediatr Infect Dis J* 30.3 (2011), pp. 227–33.

[68] Williams, E. J., Thorson, S., Maskey, M., et al. “Hospital-based surveillance of invasive pneumococcal disease among young children in urban Nepal”. *Clin Infect Dis* 48 Suppl 2 (2009), S114–22.

[69] Group, Invasive Bacterial Infection Surveillance (IBIS) and (INCLEN), International Clinical Epidemiology Network. “Prospective multicentre hospital surveillance of *Streptococcus pneumoniae* disease in India”. *Lancet* 353 (1999), pp. 1216–21.

[70] Hollingshead, S. K., Becker, R., and Briles, D. E. “Diversity of PspA: mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*”. *Infect Immun* 68 (2000), pp. 5889–900.

[71] Dave, S., Carmicle, S., Hammerschmidt, S., et al. “Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H”. *J Immunol* 173 (2004), pp. 471–7.

[72] Gilbert, R. J., Jiménez, J. L., Chen, S., et al. “Two structural transitions in membrane pore formation by pneumolysin, the pore-forming toxin of *Streptococcus pneumoniae*”. *Cell* 97 (1999), pp. 647–55.

[73] Kilian, M., Reinholdt, J., Lomholt, H., et al. “Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence”. *Acta Path Microbiol Immunol Scand* 104 (1996), pp. 321–38.

[74] Wu, Z., Deloria-Knoll, M., Hammitt, L. L., et al. “Partially latent class models for case-control studies of childhood pneumonia aetiology”. *Appl Statist* 65.1 (2016), pp. 97–114.

[75] Gilani, Z., Kwong, Y. D., Levine, O. S., et al. “A literature review and survey of childhood pneumonia etiology studies: 2000-2010”. *Clin Infect Dis* 54 Suppl 2 (2012), S102–8.

[76] Scott, J. A. G. “The global epidemiology of childhood pneumonia 20 years on”. *Bulletin of the World Health Organization* 86.6 (2008), pp. 494–496C.

[77] Deen, J., Seidlein, L. von, Andersen, F., et al. “Community-acquired bacterial bloodstream infections in developing countries in south and southeast Asia: a systematic review”. *Lancet Inf Dis* 12.6 (2012), pp. 480–487.

[78] Reddy, E. A., Shaw, A. V., and Crump, J. A. “Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis”. *Lancet* 10 (2010), pp. 417–32.

[79] Prasad, N., Murdoch, D. R., Reyburn, H., et al. “Etiology of Severe Febrile Illness in Low- and Middle-Income Countries: A Systematic Review”. *PLoS One* 10.6 (2015), e0127962.

[80] Mulholland, E. K. “Use of vaccine trials to estimate burden of disease”. *J Health Popul Nutr* 22.3 (2004), pp. 257–67.

[81] Berkley, J. A., Lowe, B. S., Mwangi, I., et al. “Bacteremia among children admitted to a rural hospital in Kenya”. *N Engl J Med* 352 (2005), pp. 39–47.

[82] Feikin, D. R., Fu, W., Park, D. E., et al. “Is Higher Viral Load in the Upper Respiratory Tract Associated With Severe Pneumonia? Findings From the PERCH Study”. *Clin Infect Dis* 64.S3 (2017), S337–S346.

[83] Baggett, H. C., Watson, N. L., Deloria Knoll, M., et al. “Density of Upper Respiratory Colonization With *Streptococcus pneumoniae* and Its Role in the Diagnosis of Pneumococcal Pneumonia Among Children Aged <5 Years in the PERCH Study”. *Clin Infect Dis* 64.S3 (2017), S317–S327.

[84] Park, D. E., Baggett, H. C., Howie, S. R. C., et al. “Colonization Density of the Upper Respiratory Tract as a Predictor of Pneumonia-*Haemophilus influenzae*, *Moraxella catarrhalis*, *Staphylococcus aureus*, and *Pneumocystis jirovecii*”. *Clin Infect Dis* 64.S3 (2017), S328–S336.

[85] Murdoch, David R, Morpeth, Susan C, Hammitt, Laura L, et al. “Microscopic analysis and quality assessment of induced sputum from children with pneumonia in the PERCH Study”. *Clinical infectious diseases* 64.suppl_3 (2017), S271–S279.

[86] Thea, D. M., Seidenberg, P., Park, D. E., et al. “Limited Utility of Polymerase Chain Reaction in Induced Sputum Specimens for Determining the Causes of Childhood Pneumonia in Resource-Poor Settings: Findings From the Pneumonia Etiology Research for Child Health (PERCH) Study”. *Clin Infect Dis* 64.S3 (2017), S289–S300.

[87] Higdon, M. M., Le, T., O’Brien, K. L., et al. “Association of C-Reactive Protein With Bacterial and Respiratory Syncytial Virus-Associated Pneumonia Among Children Aged <5 Years in the PERCH Study”. *Clin Infect Dis* 64.S3 (2017), S378–S386.

[88] Adegbola, R., Obaro, S. K., Biney, E., et al. “Evaluation of Binax NOW *Streptococcus pneumoniae* urinary antigen test in children in a community with a high rate of pneumococcus”. *Pediatr Infect Dis J* 20.7 (2001), pp. 718–9.

[89] Turner, P., Turner, C., Kaewcharernnet, N., et al. “A prospective study of urinary pneumococcal antigen detection in healthy Karen mothers with high rates of pneumococcal nasopharyngeal carriage”. *BMC Infect Dis* 11 (2011), p. 108.

[90] Moisi, J. C., Saha, S. K., Falade, A. G., et al. “Enhanced diagnosis of pneumococcal meningitis with use of the Binax NOW immunochromatographic test of *Streptococcus pneumoniae* antigen: a multisite study”. *Clin Infect Dis* 48 Suppl 2 (2009), S49–56.

[91] Deloria-Knoll, M., Moisi, J. C., Muhib, F. B., et al. “Standardizing surveillance of pneumococcal disease”. *Clin Infect Dis* 48 Suppl 2 (2009), S37–48.

[92] Azzari, C., Cortimiglia, M., Moriondo, M., et al. “Pneumococcal DNA is not detectable in the blood of healthy carrier children by real-time PCR targeting the lytA gene”. *J Med Microbiol* 60.Pt 6 (2011), pp. 710–4.

[93] 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* 51.9 (2010), pp. 1042–9.

[94] WHO. *Integrated Management of Childhood Illness (IMCI): Chart Booklet*. Geneva, Switzerland: WHO, 2014.

[95] Andrade, D. C., Borges, I. C., Adrian, P. V., et al. “Effect of Pneumococcal Conjugate Vaccine on the Natural Antibodies and Antibody Responses Against Protein Antigens From *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis* in Children With Community-acquired Pneumonia”. *Pediatr Infect Dis J* 35.6 (2016), pp. 683–9.

[96] Elberse, K. E., Tcherniaeva, I., Berbers, G. A., et al. “Optimization and application of a multiplex bead-based assay to quantify serotype-specific IgG against *Streptococcus pneumoniae* polysaccharides: response to the booster vaccine after immunization with the pneumococcal 7-valent conjugate vaccine”. *Clin Vaccine Immunol* 17.4 (2010), pp. 674–82.

[97] Dagan, R., Givon-Lavi, N., Fraser, D., et al. “Serum serotype-specific pneumococcal anticapsular immunoglobulin G concentrations after immunization with 9-valent conjugate pneumococcal vaccine correlate with nasopharyngeal acquisition of pneumococcus”. *J Infect Dis* 192 (2005), pp. 367–76.

[98] Siber, G. R., Chang, I., Baker, S., et al. “Estimating the protective concentration of anti-pneumococcal capsular polysaccharide antibodies”. *Vaccine* 25.19 (2007), pp. 3816–26.

[99] Elberse, K. E., Greeff, S. C. de, Wattimena, N., et al. “Seroprevalence of IgG antibodies against 13 vaccine *Streptococcus pneumoniae* serotypes in the Netherlands”. *Vaccine* 29.5 (2011), pp. 1029–35.

[100] Turner, P., Turner, C., Green, N., et al. “Serum antibody responses to pneumococcal colonization in the first 2 years of life: results from an SE Asian longitudinal cohort study”. *Clin Microbiol Infect* 19.12 (2013), E551–8.

[101] Andrews, N. J., Waight, P. A., Burbidge, P., et al. “Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study”. *Lancet Infect Dis* 14.9 (2014), pp. 839–846.

[102] Carter, M. J., Blohmke, C. J., and Pollard, A. J. “Immunological correlates of vaccine-mediated protection”. *The Vaccine Book, Second Edition*. Ed. by Bloom, B. R. and Lambert, P. H. London: Elsevier, 2016.

[103] Juvén, T., Mertsola, J., Waris, M., et al. “Etiology of community-acquired pneumonia in 254 hospitalized children”. *Pediatr Infect Dis J* 19.4 (2000), pp. 293–8.

[104] Nohynek, H., Eskola, J., Kleemola, M., et al. “Bacterial antibody assays in the diagnosis of acute lower respiratory tract infection in children”. *Pediatr Infect Dis J* 14.6 (1995), pp. 478–84.

[105] Mäkelä, M. J., Puhakka, T., Ruuskanen, O., et al. “Viruses and bacteria in the etiology of the common cold”. *J Clin Microbiol* 36.2 (1998), pp. 539–542.

[106] Andrade, D. C., Borges, I. C., Ivaska, L., et al. “Serological diagnosis of pneumococcal infection in children with pneumonia using protein antigens: A study of cut-offs with positive and negative controls”. *J Immunol Methods* 433 (2016), pp. 31–7.

[107] Tuerlinckx, D., Smet, J., De Schutter, I., et al. “Evaluation of a WHO-validated serotype-specific serological assay for the diagnosis of pneumococcal etiology in children with community-acquired pneumonia”. *Pediatr Infect Dis J* 32 (2013), e277–e284.

[108] Esposito, S., Bosis, S., Cavagna, R., et al. “Characteristics of *Streptococcus pneumoniae* and atypical bacterial infections in children 2?5 years of age with community-acquired pneumonia”. *Clin Infect Dis* 35 (2002), pp. 1345–52.

[109] Toikka, P., Irjala, K., Juvén, T., et al. “Serum procalcitonin, C-reactive protein and interleukin-6 for distinguishing bacterial and viral pneumonia in children”. *Pediatr Infect Dis J* 19 (2000), pp. 598–602.

[110] Virkki, R., Juven, T., Rikalainen, H., et al. “Differentiation of bacterial and viral pneumonia in children”. *Thorax* 57 (2001), pp. 438–441.

[111] Lubell, Y., Blacksell, S. D., Dunachie, S., et al. “Performance of C-reactive protein and procalcitonin to distinguish viral from bacterial and malarial causes of fever in Southeast Asia”. *BMC Infect Dis* 15.1 (2015), p. 511.

[112] Herberg, J. A., Kaforou, M., Wright, V. J., et al. “Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children”. *J Am Med Assoc* 316.8 (2016), pp. 835–45.

[113] Hammitt, L. L., Kazungu, S., Morpeth, S. C., et al. “A preliminary study of pneumonia etiology among hospitalized children in Kenya”. *Clin Infect Dis* 54 Suppl 2 (2012), S190–9.

[114] Ideh, R. C., Howie, S. R., Ebruke, B., et al. “Transthoracic lung aspiration for the aetiological diagnosis of pneumonia: 25 years of experience from The Gambia”. *Int J Tuberc Lung Dis* 15.6 (2011), pp. 729–35.

[115] Howie, S. R., Morris, G. A., Tokarz, R., et al. “Etiology of severe childhood pneumonia in the Gambia, West Africa, determined by conventional and molecular microbiological analyses of lung and pleural aspirate samples”. *Clin Infect Dis* 59.5 (2014), pp. 682–5.

[116] Feikin, D. R., Scott, J. A., and Gessner, B. D. “Use of vaccines as probes to define disease burden”. *Lancet* (383 2014), pp. 1762–70.

[117] Black, S. B., Shinefield, H. R., Ling, S., et al. “Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than five years of age for prevention of pneumonia”. *Pediatr Infect Dis J* 21.9 (2002), pp. 810–5.

[118] O’Brien, K. L., Moulton, L. H., Reid, R., et al. “Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial”. *Lancet* 362.9381 (2003), pp. 355–61.

[119] Andrade, D. C., Borges, I. C., Ivaska, L., et al. “Serological diagnosis of pneumococcal infection in children with pneumonia using protein antigens: A study of cut-offs with positive and negative controls”. *J Immunol Methods* 433 (2016), pp. 31–7.

[120] Klifka, M. K. and Ahmed, R. “Limiting dilution analysis of virus-specific memory B cells by an ELISpot assay”. *J Immunol Methods* 199 (1996)), pp. 37–46.

[121] Carter, Michael J., Mitchell, Ruth M., Meyer Sauteur, Patrick M., et al. “The antibody-secreting cell response to infection: kinetics and clinical applications”. *Frontiers in Immunology* 8 (2017).

[122] Rajewsky, J. “Clonal selection and learning in the antibody system”. *Nature* 381 (1996), pp. 751–8.

[123] McIntosh, E. D., Broker, M., Wassil, J., et al. “Serum bactericidal antibody assays - The role of complement in infection and immunity”. *Vaccine* (2015).

[124] Dang, V. D., Hilgenberg, E., Ries, S., et al. “From the regulatory functions of B cells to the identification of cytokine-producing plasma cell subsets”. *Curr Opin Immunol* 28 (2014), pp. 77–83.

[125] Bermejo, D. A., Jackson, S. W., Gorosito-Serran, M., et al. “Trypanosoma cruzi trans-sialidase initiates a program independent of the transcription factors RORgammat and Ahr that leads to IL-17 production by activated B cells”. *Nat Immunol* 14.5 (2013), pp. 514–22.

[126] Chavele, K. M., Merry, E., and Ehrenstein, M. R. “Cutting edge: circulating plasmablasts induce the differentiation of human T follicular helper cells via IL-6 production”. *J Immunol* 194.6 (2015), pp. 2482–5.

[127] Fillatreau, S., Sweenie, C. H., McGeachy, M. J., et al. “B cells regulate autoimmunity by provision of IL-10”. *Nat Immunol* 3.10 (2002), pp. 944–50.

[128] Shen, P., Roch, T., Lampropoulou, V., et al. “IL-35-producing B cells are critical regulators of immunity during autoimmune and infectious diseases”. *Nature* 507.7492 (2014), pp. 366–70.

[129] Duffy, K. R., Wellard, C. J., Markham, J. F., et al. “Activation-induced B cell fates are selected by intracellular stochastic competition”. *Science* 335.6066 (2012), pp. 338–41.

[130] Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M., et al. “The generation of antibody-secreting plasma cells”. *Nat Rev Immunol* 15.3 (2015), pp. 160–71.

[131] Tarlinton, D. and Good-Jacobson, K. “Diversity among memory B cells: origin, consequences, and utility”. *Science* 341.6151 (2013), pp. 1205–11.

[132] Shen, P., Lampropoulou, V., Stervbo, U., et al. “Intrinsic toll-like receptor signalling drives regulatory function in B cells”. *Front Biosci* 5 (2013), pp. 78–86.

[133] Hua, Z. and Hou, B. “TLR signaling in B-cell development and activation”. *Cell Mol Immunol* 10.2 (2013), pp. 103–6.

[134] Crotty, S. “T follicular helper cell differentiation, function, and roles in disease”. *Immunity* 41.4 (2014), pp. 529–42.

[135] Fernandez, J., Levine, O. S., Sanchez, J., et al. “Prevention of *Haemophilus influenzae* type B colonization by vaccination: correlation with serum anti-capsular IgG concentration”. *J Infect Dis* 182 (2000), pp. 1553–6.

[136] Consortium, MenAfriCar. “Meningococcal carriage in the African meningitis belt”. *Trop Med Int Health* 18.8 (2013), pp. 968–978.

[137] Clutterbuck, E. A., Salt, P., Oh, S., et al. “The kinetics and phenotype of the human B-cell response following immunization with a heptavalent pneumococcal-CRM conjugate vaccine”. *Immunology* 119.3 (2006), pp. 328–37.

[138] Chang, H. S. and Sack, D. A. “Development of a novel in vitro assay (ALS assay) for evaluation of vaccine-induced antibody secretion from circulating mucosal lymphocytes”. *Clin Diagn Lab Immunol* 8.3 (2001), pp. 482–8.

[139] Carpenter, C. M., Hall, E. R., Randall, R., et al. “Comparison of the antibody in lymphocyte supernatant (ALS) and ELISPOT assays for detection of mucosal immune responses to antigens of enterotoxigenic Escherichia coli in challenged and vaccinated volunteers”. *Vaccine* 24.18 (2006), pp. 3709–18.

[140] Qadri, F., Ahmed, T., Ahmed, F., et al. “Mucosal and systemic immune responses in patients with diarrhea due to CS6-expressing enterotoxigenic Escherichia coli”. *Infect Immun* 75.5 (2007), pp. 2269–74.

[141] Feller, A. J., McKenzie, R., Taylor, D. N., et al. “Comparative evaluation of the antibody in lymphocyte supernatant (ALS) and enzyme-linked immunospot (ELISPOT) assays for measuring mucosal immune responses to *Shigella* antigens”. *Vaccine* 29.47 (2011), pp. 8487–9.

[142] Qadri, F., Ryan, E. T., Faruque, A. S. G., et al. “Antigen-specific immunoglobulin A antibodies secreted from circulating B cells are an effective marker for recent local immune responses in patients with cholera: comparison to antibody-secreting cell responses and other immunological markers”. *Infect Immun* 71.8 (2003), pp. 4808–14.

[143] Leung, D. T., Rahman, M. A., Mohasin, M., et al. “Comparison of memory B cell, antibody-secreting cell, and plasma antibody responses in young children, older children, and adults with infection caused by Vibrio cholerae O1 El Tor Ogawa in Bangladesh”. *Clin Vaccine Immunol* 18.8 (2011), pp. 1317–25.

[144] Rahman, A., Rashu, R., Bhuiyan, T. R., et al. “Antibody-secreting cell responses after Vibrio cholerae O1 infection and oral cholera vaccination in adults in Bangladesh”. *Clin Vaccine Immunol* 20.10 (2013), pp. 1592–8.

[145] Sheikh, A., Bhuiyan, M. S., Khanam, F., et al. “*Salmonella enterica* serovar Typhi-specific immunoglobulin A antibody responses in plasma and antibody in lymphocyte supernatant specimens in Bangladeshi patients with suspected typhoid fever”. *Clin Vaccine Immunol* 16.11 (2009), pp. 1587–94.

[146] Khanam, F., Sheikh, A., Sayeed, M. A., et al. “Evaluation of a typhoid/paratyphoid diagnostic assay (TPTest) detecting anti-Salmonella IgA in secretions of peripheral blood lymphocytes in patients in Dhaka, Bangladesh”. *PLOS Negl Trop Dis* 7.7 (2013), e2316.

[147] Raqib, R., Rahman, J., Kamaluddin, A. K. M., et al. “Rapid diagnosis of active tuberculosis by detecting antibodies from lymphocyte secretions”. *J Infect Dis* 188 (2003), pp. 364–70.

[148] Raqib, R., Mondal, D., Karim, M. A., et al. “Detection of antibodies secreted from circulating Mycobacterium tuberculosis-specific plasma cells in the diagnosis of pediatric tuberculosis”. *Clin Vaccine Immunol* 16.4 (2009), pp. 521–7.

[149] Rekha, R. S., Kamal, S. M. M., Anderson, P., et al. “Validation of the ALS assay in adult patients with culture confirmed pulmonary tuberculosis”. *PLOS One* 6.1 (2011), e16425.

[150] Ashenafi, S., Aderaye, G., Zewdie, M., et al. “BCG-specific IgG-secreting peripheral plasmablasts as a potential biomarker of active tuberculosis in HIV negative and HIV positive patients”. *Thorax* 68.3 (2013), pp. 269–76.

[151] Chisti, M. J., Salam, M. A., Raqib, R., et al. “Validity of antibodies in lymphocyte supernatant in diagnosing tuberculosis in severely malnourished children presenting with pneumonia”. *PLOS One* 10.5 (2015), e0126863.

[152] Jiao, J., Wang, M. S., Yang, X. G., et al. “Evaluation of ALS assay of TB-SA for diagnosis of pulmonary tuberculosis”. *J Immunoassay Immunochem* 36.2 (2015), pp. 119–27.

[153] Palkola, N. V., Pakkanen, S. H., Kantele, J. M., et al. “Pathogen-specific circulating plasmablasts in patients with pneumonia”. *PLoS One* 7.3 (2012), e34334.

[154] Palkola, N. V., Blomgren, K., Pakkanen, S. H., et al. “Immune Defense in Upper Airways: A Single-Cell Study of Pathogen-Specific Plasmablasts and Their Migratory Potentials in Acute Sinusitis and Tonsillitis”. *PLoS One* 11.4 (2016), e0154594.

[155] Kantele, A., Papunen, R., Virtanen, E., et al. “Antibody-secreting cells in acute urinary tract infection as indicators of local immune response”. *J Infect Dis* 169 (1994), pp. 1023–8.

[156] Kantele, A., Möttönen, T., Ala-Kaila, K., et al. “P fimbria-specific B cell responses in patients with urinary tract infection”. *J Infect Dis* 188.1885-91 (2003).

[157] Kantele, A., Palkola, N., Arvilommi, H., et al. “Local immune response to upper urinary tract infections in children”. *Clin Vaccine Immunol* 15.3 (2008), pp. 412–7.

[158] Huang, K. Y., Li, C. K., Clutterbuck, E., et al. “Virus-specific antibody secreting cell, memory B-cell, and sero-antibody responses in the human influenza challenge model”. *J Infect Dis* 209.9 (2014), pp. 1354–61.

[159] Habibi, M. S., Jozwik, A., Makris, S., et al. “Impaired antibody-mediated protection and defective IgA B-cell memory in experimental infection of adults with respiratory syncytial virus”. *Am J Respir Crit Care Med* 191.9 (2015), pp. 1040–9.

[160] Lee, F. E., Falsey, A. R., Halliley, J. L., et al. “Circulating antibody-secreting cells during acute respiratory syncytial virus infection in adults”. *J Infect Dis* 202.11 (2010), pp. 1659–66.

[161] Lee, F. E., Halliley, J. L., Walsh, E. E., et al. “Circulating human antibody-secreting cells during vaccinations and respiratory viral infections are characterized by high specificity and lack of bystander effect”. *J Immunol* 186.9 (2011), pp. 5514–21.

[162] Wrammert, J., Onlamoon, N., Akondy, R. S., et al. “Rapid and massive virus-specific plasmablast responses during acute dengue virus infection in humans”. *J Virol* 86.6 (2012), pp. 2911–8.

[163] Garcia-Bates, T. M., Cordeiro, M. T., Nascimento, E. J., et al. “Association between magnitude of the virus-specific plasmablast response and disease severity in dengue patients”. *J Immunol* 190.1 (2013), pp. 80–7.

[164] Latner, D. R., McGrew, M., Williams, N., et al. “Enzyme-linked immunospot assay detection of mumps-specific antibody-secreting B cells as an alternative method of laboratory diagnosis”. *Clin Vaccine Immunol* 18.1 (2011), pp. 35–42.

[165] Huang, K. Y., Lin, J. J., Chiu, C. H., et al. “A potent virus-specific antibody-secreting cell response to acute enterovirus 71 infection in children”. *J Infect Dis* 212.5 (2015), pp. 808–17.

[166] Elias, S. C., Choudhary, P., Cassan, S. C. de, et al. “Analysis of human B-cell responses following ChAd63-MVA MSP1 and AMA1 immunization and controlled malaria infection”. *Immunology* 141.4 (2014), pp. 628–44.

[167] Zhang, Q., Bernatoniene, J., Bagrade, L., et al. “Serum and mucosal antibody responses to pneumococcal protein antigens in children: relationships with carriage status”. *Eur J Immunol* 36.1 (2006), pp. 46–57.

[168] Palkola, N. V., Pakkanen, S. H., Kantele, J. M., et al. “Differences in Homing Potentials of *Streptococcus pneumoniae*-Specific Plasmablasts in Pneumococcal Pneumonia and After Pneumococcal Polysaccharide and Pneumococcal Conjugate Vaccinations”. *J Infect Dis* 212.8 (2015), pp. 1279–87.

[169] Mejías, A., Suarez, N. M., and Ramilo, O. “Detecting specific infections in children through host responses: a paradigm shift”. *Curr Opin Infect Dis* 27.3 (2014), pp. 228–35.

[170] Barabasi, A. L., Gulbahce, N., and Loscalzo, J. “Network medicine: a network-based approach to human disease”. *Nat Rev Genet* 12.1 (2011), pp. 56–68.

[171] Davenport, Emma E., Burnham, Katie L., Radhakrishnan, Jayachandran, et al. “Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study”. *Lancet Resp Med* 4 (2016), pp. 259–71.

[172] Burnham, K. L., Davenport, E. E., Radhakrishnan, J., et al. “Shared and distinct aspects of the sepsis transcriptomic response to fecal peritonitis and pneumonia”. *Am J Respir Crit Care Med* (2016).

[173] Hoehn, K. B., Fowler, A., Lunter, G., et al. “The diversity and molecular evolution of B cell receptors during infection”. *Mol Biol Evol* (2016).

[174] Galson, J. D., Trück, J., Fowler, A., et al. “Analysis of B cell repertoire dynamics following hepatitis B vaccination in humans, and enrichment of vaccine-specific antibody sequences”. *EBioMedicine* 2.12 (2015), pp. 2070–9.

[175] Trück, J., Ramasamy, M. N., Galson, J. D., et al. “Identification of antigen-specific B cell receptor sequences using public repertoire analysis”. *J Immunol* 194.1 (2015), pp. 252–61.

[176] Schulze, A. and Downward, J. “Navigating gene expression using microarrays a technology review”. *Nat Cell Biol* 3 (2001), E190.

[177] Zhang, W., Yu, Y., Hertwig, F., et al. “Comparison of RNA-seq and microarray-based models for clinical endpoint prediction”. *Genome Biol* 16 (2015), p. 133.

[178] Langmead, B. and Salzberg, S. “Fast gapped-read alignment with Bowtie2”. *Nature Methods* 9 (2012), pp. 357–9.

[179] Anders, S., Pyl, P. T., and Huber, W. “HTSeq–a python framework to work with high-throughput sequencing data.” *Bioinformatics* 31 (2015), pp. 166–9.

[180] Williams, C. R., Baccarella, A., Parrish, J. Z., et al. “Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq”. *BMC Bioinformatics* 18.1 (2017), p. 38.

[181] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. *Bioinformatics* 26.1 (2010), pp. 139–40.

[182] Ramilo, O. and Meijías, A. “Shifting the paradigm: host gene signatures for diagnosis of infectious diseases”. *Cell Host Microbe* 6 (2009), pp. 199–200.

[183] Mahajan, P., Kuppermann, N., Mejias, A., et al. “Association of RNA Biosignatures With Bacterial Infections in Febrile Infants Aged 60 Days or Younger”. *JAMA* 316.8 (2016), pp. 846–57.

[184] Kafarou, M., Herberg, J. A., Wright, V. J., et al. “Diagnosis of bacterial infection using a 2-transcript host RNA signature in febrile infants 60 days or younger”. *JAMA* 317.15 (2017), pp. 1577–1578.

[185] Ramilo, O., Mejias, A., Mahajan, P., et al. “RNA signature test to distinguish bacterial from viral infection”. *J Pediatr* 182 (2017), pp. 401–404.

[186] Suarez, N. M., Bunsow, E., Falsey, A. R., et al. “Superiority of transcriptional profiling over procalcitonin for distinguishing bacterial from viral lower respiratory tract infections in hospitalized adults”. *J Infect Dis* 212.2 (2015), pp. 213–22.

[187] Tsalik, E. L., Henao, R., Nichols, M., et al. “Host gene expression classifiers diagnose acute respiratory illness aetiology”. *Sci Transl Med* 8.322 (2016), 322ra11.

[188] Sweeney, T. E., Wong, H. R., and Khatri, P. “Robust classification of bacterial and viral infections via integrated host gene expression diagnostics”. *Sci Transl Med* 8.346 (2016), 346ra91.

[189] Goh, C. and Knight, J. C. “Enhanced understanding of the host-pathogen interaction in sepsis: new opportunities for omic approaches”. *Lancet Resp Med* 5 (2017), pp. 212–23.

[190] Berry, M. P., Graham, C. M., McNab, F. W., et al. “An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis”. *Nature* 466.7309 (2010), pp. 973–7.

[191] Anderson, S. T., Kaforou, M., Brent, A. J., et al. “Diagnosis of childhood tuberculosis and host RNA expression in Africa”. *N Engl J Med* 370.18 (2014), pp. 1712–23.

[192] Pankla, R., Buddhisa, S., Berry, M., et al. “Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicaemic meliodosis”. *Genome Biol* 10.127 (2009), pp. 1–22.

[193] Wynn, J. L., Cvijanovich, N. Z., Allen, G. L., et al. “The influence of developmental age on the early transcriptomic response of children with septic shock”. *Mol Med* 17.11-12 (2011), pp. 1146–56.

[194] Parnell, G. P., McLean, A. S., Booth, D. R., et al. “A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia”. *Crit Care* 16.4 (2012), R157.

[195] Ahn, S. H., Tsalik, E. L., Cyr, D. D., et al. “Gene expression-based classifiers identify *Staphylococcus aureus* infection in mice and humans”. *PLoS One* 8.1 (2013), e48979.

[196] Cernada, M., Serna, E., Bauerl, C., et al. “Genome-wide expression profiles in very low birth weight infants with neonatal sepsis”. *Pediatr* 133 (2014), e1203–11.

[197] Smith, C. L., Dickinson, P., Forster, T., et al. “Identification of a human neonatal immune-metabolic network associated with bacterial infection”. *Nat Commun* 14.5 (2014), p. 4649.

[198] Ramilo, O., Allman, W., Chung, W., et al. “Gene expression patterns in blood leukocytes discriminate patients with acute infections”. *Blood* 109.5 (2007), pp. 2066–77.

[199] Zaas, A. K., Chen, M., Varkey, J., et al. “Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans”. *Cell Host Microbe* 6.3 (2009), pp. 207–17.

[200] Mejias, A., Dimo, B., Suarez, N. M., et al. “Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection”. *PLOS Med* 10.11 (2013), e1001549.

[201] Rautanen, Anna, Mills, Tara C., Gordon, Anthony C., et al. “Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study”. *Lancet Respiratory Medicine* 3.1 (2015), pp. 53–60.

[202] Kwan, A., Hubank, M., Rashid, A., et al. “Transcriptional instability during evolving sepsis may limit biomarker based risk stratification”. *PLOS One* 8.3 (2013), e60501.

[203] Fairfax, B. P., Humburg, P., Makino, S., et al. “Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression”. *Science* 343 (2014), p. 1118.

[204] Statistics, Central Bureau of. *National Population and Housing Census 2011*. Kathmandu, Nepal: Government of Nepal, 2012.

[205] UNDP. *Human Development Report 2016*. New York, USA: UN Development Programme, 2016.

[206] UNDP. *Human Development Report 2014: Beyond Geography, Unlocking Human Potential*. New York, USA: UN Development Programme, 2014.

[207] Health, Ministry of, (MOHP), Population, ERA, New, et al. *Nepal Demographic and Health Survey 2011*. Kathmandu, Nepal: Ministry of Health, Population, New ERA, and ICF International, 2012.

[208] Rodrigo, C. and Rajapakse, S. "Current status of HIV/AIDS in South Asia". *J Glob Infect Dis* 1.2 (2009), pp. 93–101.

[209] Hall, M. L., Lee, A. C., Cartwright, C., et al. "The 2015 Nepal earthquake disaster: lessons learned one year on". *Public Health* 145 (2017), pp. 39–44.

[210] Budhathoki, S. S. and Gelband, H. "Manmade earthquake: the hidden health effects of a blockade-induced fuel crisis in Nepal". *BMJ Glob Health* 1.2 (2016), e000116.

[211] Thapa, R., Mishra, S. R., and Gautam, R. "Deteriorating respiratory health in Nepal: a call for action". *Lancet Resp Med* (2017).

[212] Bellos, A., Mulholland, K., O'Brien, K. L., et al. "The burden of acute respiratory infections in crisis-affected populations: a systematic review". *Confl Health* 4 (2010), p. 3.

[213] Jusot, J. F., Neill, D. R., Waters, E. M., et al. "Airborne dust and high temperatures are risk factors for invasive bacterial disease". *J Allergy Clin Immunol* 139.3 (2017), 977–986 e2.

[214] Murray, P. R., Baron, E. J., Jorgensen, J. H., et al. *Manual of Clinical Microbiology (9th Edition Revised)*. Washington DC, USA: American Society for Microbiology, 2007.

[215] Clinical and Institute, Laboratory Standards. *Performance standards for antimicrobial susceptibility testing. Supplement M100-S22.* Wayne, PA, USA: Clinical and Laboratory Standards Institute, 2012.

[216] Satzke, C., Turner, P., Virolainen-Julkunen, A., et al. "Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group". *Vaccine* 32.1 (2013), pp. 165–79.

[217] Cherian, T. "WHO expert consultation on serotype composition of pneumococcal conjugate vaccines for use in resource-poor developing countries, 26-27 October 2006, Geneva". *Vaccine* 25.36 (2007), pp. 6557–64.

[218] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017.

[219] Biosystems, Applied. *Tempus Blood RNA Tube and Tempus Spin RNA Isolation Kit*. 2008.

[220] Group, SAM/BAM Format Specification Working. "Sequence Alignment Format Specification". *https://github.com/samtools/hts-specs* (2016).

[221] Chen, Y., Lun, A., McCarthy, D., et al. “Package ’edgeR’”. *http://bioinf.wehi.edu.au/edgeR* (2017).

[222] Robinson, M. D. and Oshlack, A. “A scaling normalization method for differential expression analysis of RNA-seq data”. *Genome Biol* 11.3 (2010), R25.

[223] Murdoch, Duncan and Chow, E. D. *ellipse: Functions for drawing ellipses and ellipse-like confidence regions*. R package version 0.3-8. 2013.

[224] Murdoch, D. J. and Chow, E. D. “A Graphical Display of Large Correlation Matrices”. *The American Statistician* 50.2 (1996), pp. 178–180.

[225] Fang, H., Knezevic, B., Burnham, K. L., et al. “XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits”. *Genome Med* 8 (1 2016), p. 129.

[226] Liu, J., Zhan, W., Kim, C. J., et al. “IL-10-producing B cells are induced early in HIV-1 infection and suppress HIV-1-specific T cell responses”. *PLoS One* 9.2 (2014), e89236.

[227] Higdon, M. M., Hammitt, L. L., Deloria Knoll, M., et al. “Should Controls With Respiratory Symptoms Be Excluded From Case-Control Studies of Pneumonia Etiology? Reflections From the PERCH Study”. *Clin Infect Dis* 64.S3 (2017), S205–S212.

[228] Enwere, G., Cheung, Y. B., Zaman, S. M., et al. “Epidemiology and clinical features of pneumonia according to radiographic findings in Gambian children”. *Trop Med Int Health* 12.11 (2007), pp. 1377–85.

[229] Murdoch, D. R., Morpeth, S. C., Hammitt, L. L., et al. “The diagnostic utility of induced sputum microscopy and culture in childhood pneumonia”. *Clin Infect Dis* 64.suppl_3 (2017), S280–S288.

[230] Williams, E. J., Lewis, J., John, T., et al. “*Haemophilus influenzae* type b carriage and novel bacterial population structure among children in urban Kathmandu, Nepal”. *J Clin Microbiol* 49.4 (2011), pp. 1323–30.

[231] Pradhan, R., Shrestha, U., Gautam, S. C., et al. “Bloodstream infection among children presenting to a general hospital outpatient clinic in urban Nepal”. *PLoS One* 7.10 (2012), e47531.

[232] Hothorn, T. and Zeileis, A. “partykit: A Modular Toolkit for Recursive Partytioning in R”. *J Machine Learning Res* 16 (2015), pp. 3905–3909.

[233] Carter, M. J., Emary, K. R., Moore, C. E., et al. “Rapid diagnostic tests for dengue virus infection in febrile cambodian children: diagnostic accuracy and incorporation into diagnostic algorithms”. *PLoS Negl Trop Dis* 9.2 (2015), e0003424.

[234] Stoesser, N., Moore, C. E., Pocock, J. M., et al. “Pediatric bloodstream infections in Cambodia, 2007 to 2011”. *Pediatr Infect Dis J* 32.7 (2013), e272–6.

[235] Prasad, N., Murdoch, D. R., Reyburn, H., et al. “Etiology of Severe Febrile Illness in Low- and Middle-Income Countries: A Systematic Review”. *PLoS One* 10.6 (2015), e0127962.

[236] Mollendorf, C. von, Cohen, C., Gouveia, L. de, et al. “Risk Factors for Invasive Pneumococcal Disease among Children less Than 5 Years of Age in a High HIV-Prevalence Setting, South Africa, 2010 to 2012”. *Pediatr Infect Dis J* (2014).

[237] Fancourt, N., Deloria Knoll, M., Baggett, H. C., et al. “Chest Radiograph Findings in Childhood Pneumonia Cases From the Multisite PERCH Study”. *Clin Infect Dis* 64.S3 (2017), S262–S270.

[238] Singh, J., Sundaresan, S., Manoharan, A., et al. “Serotype distribution and antimicrobial susceptibility pattern in children $\leq$5 years with invasive pneumococcal disease in India – a systematic review”. *Vaccine* (2017).

[239] Balsells, E., Guillot, L., Nair, H., et al. “Serotype distribution of *Streptococcus pneumoniae* causing invasive disease in children in the post-PCV era: A systematic review and meta-analysis”. *PLoS One* 12.5 (2017), e0177113.

[240] Hammitt, L. L., Akech, D. O., Morpeth, S. C., 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”. *Lancet Global Health* 2.7 (2014), e397–e405.

[241] Lopez Bernal, J., Cummins, S., and Gasparrini, A. “Interrupted time series regression for the evaluation of public health interventions: a tutorial”. *Int J Epidemiol* (2016).

[242] Bhaskaran, K., Gasparrini, A., Hajat, S., et al. “Time series regression studies in environmental epidemiology”. *Int J Epidemiol* 42.4 (2013), pp. 1187–95.

[243] Rückinger, S., Linden, M. van der, Reinert, R. R., et al. “Efficacy of 7-valent pneumococcal conjugate vaccination in Germany: An analysis using the indirect cohort method”. *Vaccine* 28.31 (2010), pp. 5012–6.

[244] Andrews, N., Waight, P. A., Borrow, R., et al. “Using the indirect cohort design to estimate the effectiveness of the seven valent pneumococcal conjugate vaccine in England and Wales”. *PLoS One* 6.12 (2011), e28435.

[245] Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[246] Selwyn, B. J. and BOSTID Researchers, the Coordinated Data Group of. “The epidemiology of acute respiratory tract infection in young children: comparison of findings from several developing countries”. *Rev Inf Dis* 12.8 (1990), S870–72.

[247] Conklin, L., Loo, J. D., Kirk, J., et al. “Systematic review of the effect of pneumococcal conjugate vaccine dosing schedules on vaccine-type invasive pneumococcal disease among young children”. *Pediatr Infect Dis J* 33 Suppl 2 (2014), S109–18.

[248] Nelson, J. C., Jackson, M., Yu, O., et al. “Impact of the introduction of pneumococcal conjugate vaccine on rates of community acquired pneumonia in children and adults”. *Vaccine* 26.38 (2008), pp. 4947–54.

[249] Simonsen, L., Taylor, R. J., Schuck-Paim, C., et al. “Effect of 13-valent pneumococcal conjugate vaccine on admissions to hospital 2 years after its introduction in the USA: a time series analysis”. *Lancet Resp Med* 2.5 (2016), pp. 387–94.

[250] Hortal, M., Estevan, M., Meny, M., et al. “Impact of pneumococcal conjugate vaccines on the incidence of pneumonia in hospitalized children after five years of its introduction in Uruguay”. *PLoS One* 9.6 (2014), e98567.

[251] Sgambatti, S., Minamisava, R., Bierrenbach, A. L., et al. “Early impact of 10-valent pneumococcal conjugate vaccine in childhood pneumonia hospitalizations using primary data from an active population-based surveillance”. *Vaccine* 34.5 (2016), pp. 663–70.

[252] Jardine, A., I., Menzies R., and B., McIntyre P. “Reduction in hospitalizations for pneumonia associated with the introduction of a pneumococcal conjugate vaccination schedule without a booster dose in Australia”. *Pediatr Infect Dis* 29 (2010), pp. 607–12.

[253] Weinberger, D. M., Givon-Lavi, N., Shemer-Avni, Y., et al. “Influence of pneumococcal vaccines and respiratory syncytial virus on alveolar pneumonia, Israel”. *Emerg Infect Dis* 19.7 (2013), pp. 1084–91.

[254] Bruhn, C. A., Hetterich, S., Schuck-Paim, C., et al. “Estimating the population-level impact of vaccines using synthetic controls”. *Proc Natl Acad Sci U S A* 114.7 (2017), pp. 1524–1529.

[255] Bates, M. N., Chandyo, R. K., Valentiner-Branth, P., et al. “Acute lower respiratory infection in childhood and household fuel use in Bhaktapur, Nepal”. *Environ Health Perspect* 121.5 (2013), pp. 637–42.

[256] Smith, K. R., McCracken, J. P., Weber, M. W., et al. “Effect of reduction in household air pollution on childhood pneumonia in Guatemala (RESPIRE): a randomised controlled trial”. *Lancet* 378.9804 (2011), pp. 1717–1726.

[257] Mortimer, K., Ndamala, C. B., Naunje, A. W., et al. “A cleaner burning biomass-fuelled cookstove intervention to prevent pneumonia in children under 5 years old in rural Malawi (the Cooking and Pneumonia Study): a cluster randomised controlled trial”. *Lancet* 389.10065 (2017), pp. 167–175.

[258] Darrow, L. A., Klein, M., Flanders, W. D., et al. “Air pollution and acute respiratory infections among children 0-4 years of age: an 18-year time-series study”. *Am J Epidemiol* 180.10 (2014), pp. 968–77.

[259] Kurmi, O., Regmi, P. R., and Pant, P. R. “Implication of air pollution on health effects in Nepal: lessons from global research”. *Nepal J Epidemiol* 6.1 (2016), pp. 525–7.

[260] Hamaluba, M., Kandasamy, R., Upreti, S. R., et al. “Comparison of two-dose priming plus 9-month booster with a standard three-dose priming schedule for a ten-valent pneumococcal conjugate vaccine in Nepalese infants: a randomised, controlled, open-label, non-inferiority trial”. *Lancet Infect Dis* (2015).

[261] Band, V. I., Ibegbu, C., Kaur, S. P., et al. “Induction of human plasmablasts during infection with antibiotic-resistant nosocomial bacteria”. *J Antimicrob Chemother* 69.7 (2014), pp. 1830–3.

[262] Mitchell, R., Kelly, D. F., Pollard, A. J., et al. “Polysaccharide-specific B cell responses to vaccination in humans”. *Hum Vaccin Immunother* 10.6 (2014), pp. 1661–8.

[263] Toapanta, F. R., Simon, J. K., Barry, E. M., et al. “Gut-Homing Conventional Plasmablasts and CD27(-) Plasmablasts Elicited after a Short Time of Exposure to an Oral Live-Attenuated Shigella Vaccine Candidate in Humans”. *Front Immunol* 5 (2014), p. 374.

[264] Kyu, S. Y., Kobie, J., Yang, H., et al. “Frequencies of human influenza-specific antibody secreting cells or plasmablasts post vaccination from fresh and frozen peripheral blood mononuclear cells”. *J Immunol Methods* 340.1 (2009), pp. 42–7.

[265] Sebina, I., Biraro, I. A., Dockrell, H. M., et al. “Circulating B-lymphocytes as potential biomarkers of tuberculosis infection activity”. *PLOS One* 9.9 (2014), e106796.

[266] Kelly, D. F., Snape, M. D., Clutterbuck, E. A., et al. “CRM197-conjugated serogroup C meningococcal capsular polysaccharide, but not the native polysaccharide, induces persistent antigen-specific memory B cells”. *Blood* 108.8 (2006), pp. 2642–7.

[267] Clutterbuck, E. A., Oh, S., Hamaluba, M., et al. “Serotype-Specific and Age-Dependent Generation of Pneumococcal Polysaccharide-Specific Memory B-Cell and Antibody Responses to Immunization with a Pneumococcal Conjugate Vaccine”. *Clinical and Vaccine Immunology* 15.2 (2007), pp. 182–193.

[268] Andrade, A. L., Ternes, Y. M., Vieira, M. A., et al. “Direct effect of 10-valent conjugate pneumococcal vaccination on pneumococcal carriage in children Brazil”. *PLoS One* 9.6 (2014), e98128.

[269] Elberse, K. E., Tcherniaeva, I., Berbers, G. A., et al. “Optimization and application of a multiplex bead-based assay to quantify serotype-specific IgG against *Streptococcus pneumoniae* polysaccharides: response to the booster vaccine after immunization with the pneumococcal 7-valent conjugate vaccine”. *Clin Vaccine Immunol* 17.4 (2010), pp. 674–82.

[270] Elberse, K. E., Greeff, S. C. de, Wattimena, N., et al. “Seroprevalence of IgG antibodies against 13 vaccine *Streptococcus pneumoniae* serotypes in the Netherlands”. *Vaccine* 29.5 (2011), pp. 1029–35.

[271] Pickering, J. W., Martins, T. B., Greer, R. W., et al. “A multiplexed fluorescent microsphere immunoassay for antibodies to pneumococcal capsular polysaccharides”. *Am J Clin Pathol* 117 (2002), pp. 589–596.

[272] Lopez-Raton, Monica, Rodriguez-Alvarez, M. X., Cadarso Suarez, C., et al. “OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests”. *J Stat Software* 61.8 (2014), pp. 1–36.

[273] WHO. *Training Manual for Enzyme-Linked Immunosorbent Assay for the Quantitation of Streptococcus pneumoniae serotype specific IgG (Pn PS ELISA).* London, UK and Birmingham, Alabama, USA: UCL and University of Alabama, 2014.

[274] Mariani, M., Luzzi, E., Proietti, D., et al. “A competitive enzyme-linked immunosorbent assay for measuring the levels of serum antibody to *Haemophilus influenzae* type b”. *Clin Diagn Lab Immunol* 5(5) (1998), pp. 667–74.

[275] Guiso, N., Berbers, G., Fry, N. K., et al. “What to do and what not to do in serological diagnosis of pertussis: recommendations from EU reference laboratories”. *Eur J Clin Microbiol Infect Dis* 30 (2011), pp. 307–312.

[276] Xu, G. J., Kula, T., Xu, Q., et al. “Comprehensive serological profiling of human populations using a synthetic human virome”. *Science* 348.6239 (2015), aaa0698.

[277] Sweeney, T. E., Wong, H. R., and Khatri, P. “Robust classification of bacterial and viral infections via integrated host gene expression diagnostics”. *Sci Transl Med* 8.346 (2016), 346ra91.

[278] Mahajan, P., Kuppermann, N., Mejias, A., et al. “Association of RNA Biosignatures With Bacterial Infections in Febrile Infants Aged 60 Days or Younger”. *JAMA* 316.8 (2016), pp. 846–57.

[279] Chaussabel, D., Quinn, C., Shen, J., et al. “A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus”. *Immunity* 29.1 (2008), pp. 150–64.

[280] Steenhuijsen Piters, W. A. de, Heinonen, S., Hasrat, R., et al. “Nasopharyngeal Microbiota, Host Transcriptome, and Disease Severity in Children with Respiratory Syncytial Virus Infection”. *Am J Respir Crit Care Med* 194.9 (2016), pp. 1104–1115.

[281] Henderson, J., Hilliard, T. N., Sherriff, A., et al. “Hospitalization for RSV bronchiolitis before 12 months of age and subsequent asthma, atopy and wheeze: a longitudinal birth cohort study”. *Pediatr Allergy Immunol* 16 (2005), pp. 386–92.

[282] LeMessurier, K. S., Hacker, H., Chi, L., et al. “Type I interferon protects against pneumococcal invasive disease by inhibiting bacterial transmigration across the lung”. *PLoS Pathog* 9.11 (2013), e1003727.

[283] Okeke, I. N., Laxminarayan, R., Bhutta, Z. A., et al. “Antimicrobial resistance in developing countries. Part I: recent trends and current status”. *Lancet Infect Dis* 5.8 (2005), pp. 481–493.