

https://dspace.org/

Molecular Characterisation and Population Structure Determination of Streptococcus pneumoniae in South Africa using Whole-Genome Analysis

Lekhuleni, Cebile Natasha Nolwazi

2025
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/49956

Lekhuleni, Cebile Natasha Nolwazi . (2025). Molecular Characterisation and Population Structure Determination of Streptococcus pneumoniae in South Africa using Whole-Genome Analysis [PhD thesis, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/49956

Downloaded from a DSpace repository. https://dspace.org/

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

Molecular Characterisation and Population Structure Determination of *Streptococcus pneumoniae* in South Africa using Whole-Genome Analysis

Cebile Natasha Nolwazi Lekhuleni

717030

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Supervisors:

Dr Mignon du Plessis (PhD), Medical Scientist

Dr Kedibone Ndlangisa (PhD), Medical Scientist

Prof Anne von Gottberg (MBBCH, DTM&H, FCPath (SA) MICRO, PhD),

Clinical Microbiologist

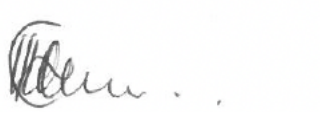
Johannesburg, 30 September 2025

Student declaration

I, Cebile Lekhuleni (student number: 717030), declare that the thesis is my own, unaided work (no plagiarism). It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

The research work described in this thesis was conducted under the supervision of Dr Mignon du Plessis, Dr Kedibone Ndlangisa and Prof. Anne von Gottberg, in the Center for Respiratory Diseases and Meningitis at the National Institute for Communicable Diseases.

Specimens and data used in this thesis were from the Group for Enteric, Respiratory and Meningeal Disease Surveillance – South Africa (GERMS-SA) programme. I (Cebile Lekhuleni) have been involved in the secondary analysis of the data arising from GERMS-SA (detailed in this thesis) which included isolate retrieval from cryopreservation, DNA extraction and preparation for whole-genome sequencing, troubleshooting, data analysis (bioinformatics and statistical analysis), data management (data cleaning, data transfer and long-term storage), and first author manuscript writing. Chapter 3 of this thesis has been published, and Chapter 4 is under peer-review. All co-authors have agreed to the use of the manuscripts as part of this thesis.



_____

30 September 2025, Johannesburg

Dedication

To my beloved mother

Nozipho Joyce Shongwe

For all your hard work and sacrifices to get me to this level

List of presentations

Lekhuleni C, Ndlangisa K, Lo SW, de Gouveia L, Kleynhans J, Quan V, Hawkins PA, McGee L, Bentley SD, Cohen C, von Gottberg A, and du Plessis M. Genomic epidemiology and antimicrobial resistance of invasive pneumococcal serotype 8 in the pre- and post-PCV13 eras in South Africa, 2005-2020. 12th International Symposium on Pneumococci & Pneumococcal Diseases (ISPPD-12). June 19-23, 2022. Toronto, Canada. **Poster presentation.**

Lekhuleni C. Genomic epidemiology and antimicrobial resistance of invasive pneumococcal serotype 8 in the pre- and post-PCV13 eras in South Africa, 2005-2020. Wits Faculty of Health Sciences Research Day September 15, 2022. Johannesburg, South Africa. **Oral presentation.**

Lekhuleni C. Impact of pneumococcal conjugate vaccines on invasive pneumococcal disease-causing lineages in South African children. 1st Public Health Alliance for Genomic Epidemiology (PHA4GE), October 30-November 1, 2023. Cape Town, South Africa. **Oral presentation.**

Lekhuleni C. Genomic surveillance strategies for vaccine-preventable bacterial diseases with a focus on AMR and vaccine design. 6th African Society for Laboratory Medicine (ASLM), December 12-15, 2023. Cape Town, South Africa. **Oral presentation.**

Lekhuleni C, Ndlangisa K, Chochua S, Metcalf BJ, Li Y, Hazelhurst S, de Gouveia L, Kleynhans J, Ismail A, Walaza S, Quan V, Meiring S, Hawkins PA, McGee L, Bentley SD, Cohen C, Lo SW, von Gottberg A, and du Plessis M. Pneumococcal lineages and mechanisms of antimicrobial resistance in the post-PCV13 era in South Africa, 2015-2020. 13th International Society of Pneumonia & Pneumococcal Diseases (ISPPD-13), March 17-20, 2024. Cape Town, South Africa. **Poster presentation.**

List of publications

Lekhuleni C, Ndlangisa K, Gladstone RA, Chochua S, Metcalf BJ, Li Y, Kleynhans J, de Gouveia L, Hazelhurst S, Ferreira AD, Skosana H, Walaza S, Quan V, Meiring S, Hawkins PA, McGee L, Bentley SD, Cohen C, Lo SW, von Gottberg A, du Plessis M. Impact of pneumococcal conjugate vaccines on invasive pneumococcal disease-causing lineages among South African children. Nature Communications. 2024 Sep 27;15(1):8401.

Lekhuleni C, Ndlangisa K, Ferreira AD, Cheng HR, Lorenz O, de Gouveia L, Quan V, Hawkins PA, McGee L, Bentley SD, Meiring S, Cohen C, Lo SW, du Plessis M, von Gottberg A. Trends in antibiotic resistant invasive pneumococcal disease caused by non-13-valent pneumococcal conjugate vaccine serotypes and lineages in South Africa. *In preparation.*

Lekhuleni C, Ndlangisa K, Ferreira AD, Cheng HR, Lorenz O, de Gouveia L, Quan V, Hawkins PA, McGee L, Bentley SD, Meiring S, Cohen C, Lo SW, von Gottberg A, du Plessis M. Clonal Expansion of Global Pneumococcal Sequence Cluster 3 within Serotype 8 after 13-Valent Pneumococcal Conjugate Vaccine Introduction, South Africa. *In preparation.*

Abstract

Pneumococcal conjugate vaccines (PCV) have effectively reduced overall invasive pneumococcal disease (IPD) and antibiotic-nonsusceptible IPD in many countries. Ongoing invasive disease due to residual vaccine-targeted serotypes (VT) and non-vaccine serotypes (NVT) after PCV introduction remains a global concern. In South Africa, seven-valent PCV (PCV7) was introduced in 2009 and was replaced by 13-valent PCV (PCV13) in 2011. We aimed to evaluate changes in IPD lineages and antibiotic-nonsusceptibility before and after PCV introduction in South Africa.

We analysed 54,199 IPD cases reported from 2005–2020 as part of the national, laboratory-based surveillance program. A randomly selected subset of approximately 4800 isolates underwent whole-genome sequencing. We determined *in silico* serotypes, sequence types, global pneumococcal sequence cluster (GPSC) lineages, and antimicrobial resistance determinants from the genomes of 4669 isolates. Vaccine periods were defined as pre-PCV (2005–2008), PCV7 (2009–2010), early-PCV13 (2011–2014), and late-PCV13 (2015–2020). Penicillin multidrug resistance (penicillin-MDR) was defined as nonsusceptibility to penicillin plus ≥2 other antibiotic classes. Other MDR was defined as nonsusceptibility to ≥3 antibiotic classes. Vaccine-type and non-vaccine-type lineages were defined as expressing ≥50% VTs and >50% NVTs, respectively, before PCV introduction. *In silico* serotype and lineage diversity trends were determined using Simpson's diversity index (SDI).

Significant incidence reductions occurred among vaccine-type lineages in the late-PCV13 period compared to the pre-PCV period. The five most common serotypes in the late-PCV13 period were NVTs 8, 12F, 15B/C, and residual VTs 19A and 19F, mainly expressed by GPSC3 (non-vaccine-type lineage), GPSC26 and GPSC56 (non-vaccine-type and vaccine-type), GPSC48 (non-vaccine-type), GPSC17 (vaccine-type), and GPSC1/21 (vaccine-types), respectively. Compared to the baseline period, significantly higher serotype diversity estimates were observed in the early-PCV13 period, with the highest estimate occurring during the PCV13 introduction year (2011, SDI 0.993, 95% confidence interval [CI] 0.991–0.995). A significant increase in lineage diversity was also observed from the PCV7 period to the early-PCV13 period (SDI: 0.954, 95% CI 0.948–0.961 vs 0.965, 0.962–0.969) supporting intervention-driven population structure perturbation. Overall antibiotic-nonsusceptible IPD declined across the study period. Residual PCV13 serotypes 19A and 19F contributed 44% (977/2236) of penicillin-cotrimoxazole or penicillin-MDR IPD in the late-PCV13 period. Among NVTs, serotypes 10A, 15A, 15B/C, 16F, 23B, and 35B were common causes (61.9%, 545/881) of penicillin-cotrimoxazole or penicillin-MDR IPD after PCV13 introduction in children and adults. Overall, NVT 8 was the most common cause of IPD after PCV13 introduction with significant incidence increases among children ≤5 years (incidence rate ratio 2.3, 95% CI 1.3–4.1) and adults >64 years (2.8, 1.1–8.3) when comparing pre-PCV versus late-PCV13 periods. NVT 8 (GPSC3, sequence type 53) IPD was predicted to continue increasing despite remaining largely susceptible to first-line treatment.

Incidence of IPD declined in all ages in South Africa after PCV introduction. Similarly, overall nonsusceptible IPD declined in the post-PCV era. Vaccine-type lineages, particularly those expressing VTs 19A (GPSC17) and 19F (GPSC21) continued to circulate after PCV introduction and were important contributors to penicillin-nonsusceptible IPD in addition to

some pre-existing non-vaccine-type lineages. GPSC3 (NVT 8) caused substantial disease in children and adults in the post-PCV era. In 2024, South Africa switched to a lower-valency vaccine PCV10 (Pneumosil, Serum Institute of India). Given the expansion of non-vaccine-type lineages and the persistence of vaccine-type lineages that are predominantly expressing their NVT counterparts following PCV use, it is crucial to monitor the impact of this switch on the pneumococcal population structure and AMR to inform future PCV formulation considerations.

Acknowledgements

Academic acknowledgements

To my supervisors, *Dr Mignon du Plessis*, *Dr Kedibone Ndlangisa*, and *Prof Anne von Gottberg* – there are not enough words to express how grateful I am for the teachings, guidance, supervision, patience, and support you have shown me. In the academic space, I am beyond blessed to have grown under your supervision. My PhD journey has been nothing short of amazing because of you. Thank you.

Prof Anne von Gottberg and Prof Lee Harrison – Thank you for giving me the life changing opportunity to be part of SAPPHGenE. I could not have asked for a better opportunity. I would do this again in a heartbeat.

To my SAPPHGenE mentors, *Dr Sopio Chochua*, *Dr Benjamin Metcalf*, *Dr Yuan Li* – Thank you for all that you have taught me in the field of genomic surveillance and bioinformatics. The knowledge you have imparted is invaluable.

To my personal mentor, *Dr Stephanie Lo* – I have learnt so much from you. You made sure to take time out and assist with my work, giving great suggestions and making sure that I have all the help I need from the Sanger Institute. Thank you and I appreciate you.

To *Prof Cheryl Cohen*, *Dr Jackie Kleynhans*, and *Dr Sibongile Walaza* – Thank you so much for your valuable input and assistance in this journey.

Linda de Gouveia, Happy Skosana, Dineo Mogale, and Rivionia Nero – Thank you for teaching me phenotypic work and always being available when I needed assistance with the surveillance database and isolates. I appreciate your contribution.

To my SAPPHGenE buddies, *Buhle Ntozini*, *Rito Mikhari-Chauke*, *Dikeledi Kekana*, *Sibusisiwe Zulu*, *Nosihle Msomi*, and *Rutendo Mapengo* – this journey has been nothing but bliss with you all in it. We had so many special moments together and I thank you all for that.

Fahima Moosa – thank you for all your assistance, support and all the doses of positivity every morning during my thesis writing time. Those gave me the push I needed, each time.

Personal acknowledgements

Nozipho Shongwe – You were my cornerstone in all this. Thank you for being my biggest cheerleader. You are the best mom in the world.

Musawenkosi Nkosi – Thank you for all the late nights you had to pick me up from work, supporting my every decision, advice and being a shoulder to lean on. I could not have asked for a better partner.

Nontokozo Lekhuleni – Thank you for being such an amazing and supportive sister. Your words of encouragement and prayers went a long way.

Funding acknowledgment

I would like to thank the following funding body for making this project a possibility:

SAPPHGenE program: Fogarty International Center Global Infectious Disease research training grant, National Institutes of Health, to the University of Pittsburgh and National Institute for Communicable Diseases – cooperative agreement number D43TW011255.

Table of Contents

Student declaration i

Dedication ii

List of presentations iii

List of publications v

Abstract vi

Acknowledgements ix

List of Figures xvii

List of Tables xxi

List of Abbreviations xxii

Chapter One - Literature review 1

1. Introduction 1

1.1. *Streptococcus pneumoniae* 1

1.2. Pneumococcal disease development 2

1.3. Pneumococcal disease burden and epidemiology 2

1.4. Prevention of pneumococcal disease 4

1.4.1. Pneumococcal polysaccharide vaccines 4

1.4.2. Pneumococcal conjugate vaccines 5

1.4.3. Impact of pneumococcal conjugate vaccines and serotype replacement 6

1.5. Treatment of pneumococcal disease – antibiotic therapy 7

1.6. Antimicrobial resistance in *S. pneumoniae* 8

1.7. Molecular characterisation of *S. pneumoniae* 10

1.7.1. Seven-locus sequence typing 10

1.7.2. Global Pneumococcal Sequence Clusters (GPSC) 11

1.7.3. Bioinformatics tools for *S. pneumoniae* genomic analysis 13

1.7.4. Tools for phylogenetics and evolutionary inference 14

1.8. Problem statement and importance of current study 16

1.9. Study Aim and Objectives 17

Chapter Two - Methods and materials 18

2. IPD surveillance, study population and laboratory methods 18

2.1. GERMS-SA surveillance programme 18

2.2. Isolate processing and phenotypic characterisation 19

2.3. Study population and sampling strategy 21

2.4. DNA extraction and whole-genome sequencing 22

2.5. Genomic data processing and bioinformatics analysis 23

Chapter Three - Impact of pneumococcal conjugate vaccines on invasive pneumococcal disease-causing lineages among South African children 26

3. Chapter 3 reference: 26

3.1. Introduction 26

3.2. Methods and materials 27

3.2.1. National invasive pneumococcal disease surveillance and sample selection…. 27

3.2.2. Whole-genome sequencing and processing 28
3.2.3. Species-wide phylogenetic visualisation 28
3.2.4. Lineage-specific analyses 28
3.2.5. Dating and phylodynamic analysis 28
3.2.6. *In silico* antimicrobial resistance determinant analyses 29
3.2.7. Statistical analysis 30
3.3. Results 31
3.3.1. National invasive pneumococcal disease surveillance 31
3.3.2. Serotype distribution 32
3.3.3. Genomic population structure and lineage-specific incidence changes 33
3.3.4. Phylodynamic analysis 35
3.3.5. Simpson's diversity index trends 35
3.3.6. Antimicrobial resistance 36
3.4. Discussion 47
Chapter Four - Trends in antibiotic resistant invasive pneumococcal disease caused by non-13-valent pneumococcal conjugate vaccine serotypes and lineages in South Africa 53
4. Chapter 4 reference: 53
4.1. Introduction 53
4.2. Methods and materials 54
4.2.1. Invasive pneumococcal disease surveillance in South Africa 54
4.2.2. Phenotypic characterisation 54
4.2.3. Whole-genome sequencing and molecular characterisation 55
4.2.4. Statistical analysis 55
4.3. Results 56

4.3.1. Invasive pneumococcal disease surveillance, South Africa 56
4.3.2. Trends in antibiotic-nonsusceptible invasive disease 56
4.3.3. Residual PCV13 and non-PCV13 serotype antimicrobial nonsusceptibility among viable isolates 57
4.3.4. Genomic features of antibiotic-resistant non-PCV13 serotypes using sampled data 58
4.4. Discussion 73
Chapter Five - Clonal Expansion of Global Pneumococcal Sequence Cluster 3 within Serotype 8 after 13-Valent Pneumococcal Conjugate Vaccine Introduction, South Africa 79
5. Chapter 5 reference: 79
5.1. Introduction 79
5.2. Methods and materials 80
5.2.1. Invasive pneumococcal disease surveillance in South Africa 80
5.2.2. Phenotypic characterisation 80
5.2.3. Whole-genome sequencing and molecular characterisation 81
5.2.4. Phylogenetic analysis 81
5.2.5. Clonal expansion analysis 82
5.2.6. Statistical analysis 82
5.3. Results 83
5.3.1. National invasive pneumococcal disease surveillance, South Africa 83
5.3.2. Antimicrobial resistance 83
5.3.3. Serotype 8 genotypes and GPSC lineages, South Africa 84
5.3.4. Global distribution of serotype 8 genotypes and GPSC lineages 84
5.3.5. Lineage-specific analysis - GPSC3 85

5.3.6. Clonal expansion of CC53 (GPSC3) genomes from South Africa.....86

5.4. Discussion.....97

Chapter six - Conclusion.....101

References.....104

Appendices.....125

List of Figures

Figure 2.1: Map of South Africa showing GERMS-SA participating laboratories across all nine provinces. Blue circles are indicative of <10 laboratories in a region. Orange indicates ≥10 laboratories…….19

Figure 2.2: Flowchart showing total numbers of IPD cases and sequenced isolates from persons of all ages during the 2005–2020 period in South Africa……21

Figure 3.1: Single nucleotide polymorphism (SNP) maximum-likelihood phylogeny (N=3104) of invasive pneumococcal disease isolates from South African children <18 years……38

Figure 3.2: Global Pneumococcal Sequence Cluster (GPSC) lineages among invasive disease isolates from (a) children <5 years (n=2680) and (b) 5-17 years (n=424) across vaccine periods in South Africa (N=3104)……39-41

Figure 3.3: Temporal trend analysis of important vaccine-type lineages causing persistent invasive disease in the late-PCV13 period inferred using skygrowth……42

Figure 3.4: Clonal expansion dynamics of GPSC9 (a) and GPSC10 (b) inferred using CaveDive……43

Figure 3.5: Simpson's diversity index (SDI) estimates for invasive disease serotypes and lineages…….44

Figure 4.1: Incidence of reported invasive pneumococcal disease (IPD) in children <5 years (a) and individuals ≥5 years (b) in South Africa, 2005–2020……..64

Figure 4.2: Incidence of nonsusceptible invasive pneumococcal disease (IPD) in children <5 years and individuals ≥5 years in South Africa, 2005–2020……65

Figure 4.3: Antibiotic-nonsusceptible IPD by vaccine period in all ages, South Africa. a–d show nonsusceptibility trends in the pre-PCV (2005–2008), PCV7 (2009–2010), early-PCV13 (2011–2014), and late-PCV13 (2015–2020) periods……..66

Figure 4.4: Global Pneumococcal Sequence Cluster (GPSC) lineages among residual PCV13 *in silico* serotypes 19A (a) and 19F (b) in South Africa, 2005–2020……..67

Figure 4.5: Multilocus sequence typing (MLST) minimum-spanning-trees of the two most common penicillin-nonsusceptible PCV13 serotypes 19A (a) and 19F (b) in South Africa in the late-PCV13 period…….68-69

Figure 4.6: Global Pneumococcal Sequence Cluster (GPSC) lineages among the six (a-f) most common penicillin-nonsusceptible non-PCV13 serotypes in South Africa, 2005–2020……70

Figure 4.7: Multilocus sequence typing (MLST) minimum-spanning-trees of the six (a-f) most common penicillin-nonsusceptible non-PCV13 serotypes in South Africa in the late-PCV13 period. Sequence types (ST) are coloured by Global Pneumococcal Sequence Cluster (GPSC) lineages……71-72

Figure 5.1: Serotype 8 trends and characteristics, 2005–2020, South Africa. a) Incidence of serotype 8 invasive pneumococcal disease reported by year and by age group (N=2429).…90

Figure 5.2: Contextualisation of South African serotype 8 genomes…….91-92

Figure 5.3: Global distribution of GPSC3 lineage (N=749)……92

Figure 5.4: GPSC3 *cps* loci comparative analysis.……..93

Figure 5.5: GPSC3 Phandango plot of recombination detected with Gubbins, focused on the *cps* locus, across GPSC3 phylogeny, N=749…..…….94

Figure 5.6: CC53 expansion dynamics among genomes from South Africa (N=219) inferred using CaveDive……..95-96

List of Tables

Table 3.1: Changes in the prevalence of antibiotic nonsusceptibility among South African children, comparing the pre-PCV and late-PCV13 periods……45-46

Table 4.1: Characteristics of invasive pneumococcal disease cases, South Africa, 2005–2020. N=51,917 (total number of reported cases with known age), n=35,416 (total number of viable isolates with known age), n_seq=4669 (total number of sequenced isolates)……61-63

Table 5.1: Serotype 8 invasive pneumococcal disease isolates stratified by vaccine period (pre-PCV, PCV7, early-PCV13, and late-PCV13) among individuals of all ages in South Africa, 2005–2020, N=2429 (number of cases with known age), n=391 (number of sequenced isolates)……...88-89

List of Abbreviations

%	–	percentage
<	–	less than
>	–	greater than
≤	–	less than or equal to
≥	–	greater than or equal to
°C	–	degrees Celsius
µl	–	microliter
AMR	–	antimicrobial resistance
BLAST	–	Basic Local Alignment Search Tool
bp	–	base pair
CLSI	–	Clinical and Laboratory Standards Institute
CC	–	clonal complex
CDC	–	Centers for Disease Control and Prevention
CI	–	confidence interval
cps	–	capsular polysaccharide
CRDM	–	Centre for Respiratory Diseases and Meningitis
CSF	–	cerebrospinal fluid
DDD	–	defined daily dose
DLV	–	double-locus variant
DMP	–	Diagnostic Media Product
DNA	–	Deoxyribonucleic acid
ESS	–	effective sample size

GERMS-SA Surveillance-South Africa	–	Group for Enteric, Respiratory and Meningeal Disease
GPS	–	Global Pneumococcal Sequencing project
GPSC	–	global pneumococcal sequence cluster
HGT	–	horizontal gene transfer
HIV	–	human immunodeficiency virus
IPD	–	invasive pneumococcal disease
IRR	–	incidence rate ratio
LRI	–	lower respiratory infection
MCMC	–	Markov chain Monte Carlo
MDR	–	multidrug resistant
MIC	–	minimum inhibitory concentration
MLST	–	multilocus sequence typing
NHLS	–	National Health Laboratory Service
NICD	–	National Institute for Communicable Diseases
NT	–	non-serotypeable
NVT	–	non-vaccine serotype
PCV	–	pneumococcal conjugate vaccine
PCV7	–	7-valent pneumococcal conjugate vaccine
PCV10	–	10-valent pneumococcal conjugate vaccine
PCV13	–	13-valent pneumococcal conjugate vaccine
PCV15	–	15-valent pneumococcal conjugate vaccine
PCV20	–	20-valent pneumococcal conjugate vaccine
PCR	–	polymerase chain reaction
PFGE	–	pulsed-field gel electrophoresis

PMEN	–	pneumococcal molecular epidemiology network
PPV	–	pneumococcal polysaccharide vaccine
SDI	–	Simpson's index of diversity
SLV	–	single-locus variant
SNP	–	single nucleotide polymorphism
ST	–	sequence type
UK	–	United Kingdom
USA	–	United States of America
VT	–	vaccine serotype
WGS	–	whole-genome sequencing
WHO	–	World Health Organization

Chapter One - Literature review

1. Introduction

1.1. *Streptococcus pneumoniae*

The genus *Streptococcus* is diverse and consists of 159 species [1], of which some are commensals and others are human pathogens. *Streptococcus pneumoniae* (or pneumococcus) is one of the clinically relevant bacterial species that causes disease in humans. The pneumococcus is a Gram-positive, encapsulated diplococcus that is part of the human respiratory tract commensal flora and is usually associated with asymptomatic colonisation, predominantly in children [2]. *S. pneumoniae* gained historical significance due to its association with respiratory infection and substantial global burden of pneumococcal disease, especially among vulnerable populations [3].

The pneumococcus produces structurally diverse capsular polysaccharides that are immunogenic and can be phenotypically identified using specific antisera that underlie the serotyping method. By 2020, at least 100 pneumococcal serotypes were identified [4]. These serotypes can further be grouped into 46 immunologically similar serogroups [5]. The capsule of *S. pneumoniae* is composed of polysaccharide subunits that are located in the capsular polysaccharide (*cps*) operon, flanked by the *dexB* and *aliA* genes [6]. The first four genes of the *cps* operon, namely: *wzg, wzh, wzd* and *wze* (also referred to as *cpsA*, *cpsB*, *cpsC and cpsD*, respectively) are conserved in all pneumococcal serotypes (except serotype 37) and are responsible for capsule processing and regulation [6]. However, the partial or complete loss of the *cps* operon [7], *cps* gene replacement [7–9], single nucleotide

polymorphisms (SNPs) within the *wchA* (*cpsE*) gene, or sequence duplication within the *cps* operon [10], may results in non-typeable pneumococci due to the lack of a capsule.

1.2. Pneumococcal disease development

Pneumococcal colonisation is a pre-requisite for pneumococcal disease development and the polysaccharide capsule of the bacterium plays an important role in pathogenesis and inhibits complement activity and phagocytosis [4,11,12]. Colonisation (also known as carriage) has been shown to be high in certain populations with children exhibiting higher carriage rates than adults [13,14]. The progression from colonisation to pneumococcal disease occurs by the contiguous spread from the nasopharynx to cause non-invasive disease such as bronchitis, conjunctivitis, otitis media, or sinusitis [13]. Invasive pneumococcal disease (IPD) results from the bacterium entering the bloodstream causing bacteraemia and can disseminate to infect sterile sites such as the lungs (pneumonia), or meninges (meningitis) [15]. Children <5 years of age, the elderly, and the immunocompromised are at a higher risk of pneumococcal infection [16] due to their inability to effectively mediate antibody-initiated complement opsonisation which activates the classical complement pathway that protects against disease [13,17].

1.3. Pneumococcal disease burden and epidemiology

S. pneumoniae is an important cause of community-acquired pneumonia, bacteraemia and meningitis in children and adults, globally [18]. Before pneumococcal vaccine use, pneumococcal disease accounted for over one million deaths each year (2000–2003) in children ≤5 years of age, with the highest IPD burden of approximately 90% of the total global fatalities occurring in low- and middle-income countries in Asia, Latin America and

Sub-Saharan Africa [19–21]. Studies have also shown that pneumococcal disease burden is higher among younger children, the elderly, and the immunocompromised [22,23]. Approximately ten pneumococcal serogroups were responsible for the majority of paediatric IPD in different continents with serogroups 1, 6, 14, 19, and 23 being the most common in each continent and also predominantly causing antibiotic nonsusceptible IPD [24]. The same serogroups predominated in older children and adult populations. Prior to vaccine introduction, between 70%–90% of IPD in North America and Europe, 50%–85% in Oceania, Latin America, and Asia, and 40%–70% of IPD in Africa was caused by serotypes in the predominant serogroups that later became candidates for different pneumococcal vaccine formulations [25]. In the USA, serotypes that commonly caused IPD in children <5 years before vaccine introduction were serotype 4, 6B, 14, 18C, 19F, and 23F, and among the ≥5 years, serotype 4, 6B, 9V, 14, and 23F were important contributors to disease [26]. In Europe and some African countries, similar serotypes were common causes of IPD, however with serotype 1 and 5 showing increased importance in these settings [5,27].

Before pneumococcal conjugate vaccine (PCV) introduction in South Africa (2005–2008), approximately 107,600 children <5 years were hospitalised annually with pneumococcal disease [28]. Serotypes 4, 6B, 9V, 14, 18C, 19F and 23F were responsible for 58% (3849/6668) of IPD cases among children <5 years for surveillance years 2003–2008 [29]. In older children and adults, 70% of IPD was caused by the same serotypes plus 1, 3, 5, 6A, 7F, and 19A [30]. Before vaccine introduction, IPD incidence among children <1 and 1–4 years was 87 and 14 cases per 100,000 population, respectively [29]. However, South Africa has a high prevalence of HIV infection [30], and a 21-fold (95% CI, 19–24) and 34-fold (95% CI, 29–41) greater relative risk of IPD was observed in HIV-infected children compared to HIV-uninfected children among the <1 and 1–4 years age groups, respectively [29]. Additionally,

IPD burden was also high in the 25–45 year olds [22]. IPD incidence in HIV-infected individuals was estimated to be 43 times higher (52 per 100 000 vs. 1.2 per 100,000) than in HIV-uninfected persons [30].

1.4. Prevention of pneumococcal disease

1.4.1. Pneumococcal polysaccharide vaccines

The polysaccharide capsule of the pneumococcus is an important virulence determinant that is targeted for vaccine formulations due to its immunologic effect. In 1977, the first pneumococcal polysaccharide vaccine (PPV) formulated using 14 (PPV14) purified capsular polysaccharide antigens (1, 2, 3, 4, 5, 6A, 7F, 8, 9N, 12F, 18C, 19F, 23F, 25F) that were commonly causing pneumococcal disease was licenced in the USA [31]. In 1983, PPV14 was replaced by PPV23 (Pneumovax23™, Merck) that targeted an additional 11 serotypes (6B, 9V, 10A, 11A, 14, 15B, 17F, 19A, 20, 22F, 33F), excluding serotype 6A and 25F [31,32]. PPVs elicit T-cell independent antibody response by the direct stimulation of immune response in B-cells [33]. However, due to poor immune response in children younger than 2 years as a result of immature immune systems that cannot mount a strong or lasting immune response to T-cell independent antigens, vaccination with PPV was recommended for older children, adults, and immunocompromised individuals [3,31,34,35]. In South Africa, PPV23 first became available in 1992 (Pneumovax® MSD) and then 1998 (Imovax Pneumovax 23® Aventis) and was recommended for use in adults aged 19–64 years with underlying comorbidities or immunocompromising conditions, and for all persons ≥65 years [36,37].

1.4.2. Pneumococcal conjugate vaccines

Third-generation PCVs were developed by covalently conjugating polysaccharide antigens to immunogenic carrier proteins such as the non-toxic mutant of diphtheria toxin (CRM197) and other carriers such as protein D of non-typeable *Haemophilus influenzae* for increased immunogenicity [31,32,38,39]. Not only does the presence of carrier proteins render PCV highly immunogenic, the protective T-cell-dependent immune response spans across all ages, more notably children less than 2 years. Unlike PPV, PCV has the additional benefit of indirect effect through the indirect protection of unvaccinated persons by reducing the overall circulation of the pneumococcus (as mucosal immunity decreases acquisition and carriage of vaccine serotypes in vaccinated infants and children) among susceptible populations, especially the elderly [40]. PCV7 (Pfizer, NY, USA) targeting serotype 4, 6B, 9V, 14, 18C, 19F, and 23F was first introduced into childhood immunisation in the USA in 2000 and targeted serotypes that were the most common cause of paediatric invasive disease, responsible for approximately 80% of IPD in children younger than 5 years of age [24,40,41]. Additionally, these serotypes contributed to high carriage rates and antimicrobial resistance (AMR) [42].

Several higher-valency PCV formulations such as 10-valent PCV (PCV10, GSK, Rixensart, Belgium) targeting PCV7 serotypes plus serotypes 1, 5, and 7F; PCV13 (Pfizer, NY, USA) targeting PCV10 serotypes plus serotypes 3, 6A, and 19A; 15-valent PCV (PCV15, Merck, NJ, USA) which targets PCV13 serotypes plus serotypes 22F and 33F, and 20-valent PCV (PCV20, Pfizer, NY, USA) targeting PCV15 serotypes plus serotypes 8, 10A, 11A, 12F, and 15B/C have been developed to improve vaccine coverage. PCV15 and PCV20 are currently recommended for use in children and adults in the USA and some European countries [43,44]. A 24-valent PCV (PCV24) that targets PCV20 serotypes plus serotypes 2, 9N, 17F,

and 20B is currently under development. A novel, more affordable PCV10 (Pneumosil, Serum Institute of India, Pune, India) that targets PCV13 serotypes excluding serotypes 3, 4, and 18C was developed by the Serum Institute of India [45]. In South Africa, PCV7 was introduced into the routine immunisation programme in 2009 and was administered in a 2 + 1 schedule at 6, 14, and 40 weeks of age, and was replaced by PCV13 in 2011 [22,46]. The more cost-effective PCV10 (Pneumosil, Serum Institute of India) replaced PCV13 in 2024 using the same schedule (45).

1.4.3. Impact of pneumococcal conjugate vaccines and serotype replacement

PCVs have substantially reduced invasive disease caused by vaccine-targeted serotypes (VT) in many countries [47–49]. Disease reduction in unvaccinated children and adults due to indirect protection was also reported [50,51]. In South Africa, surveillance for laboratory-confirmed IPD was initiated in 1999 and is ongoing [22,52]. A decline of 69% in IPD (54.8 to 17.0 cases per 100,000 population) among children <2 years was observed in South Africa from the baseline pre-PCV period (2005–2008) to 2012, including a decline of 89% (32.1 to 3.4 cases per 100,000 person-years) in disease caused by PCV7 serotypes [22]. Sustained IPD incidence reduction of 32.4 to 1.9 and 14.0 to 1.0 cases per 100,000 population due to PCV7 and PCV13-unique serotypes, respectively, were reported by 2019 in the same age group [47].

However, despite the success of PCVs IPD remains an important health concern due to ongoing disease caused by previously circulating non-vaccine serotypes (NVT). Several countries have observed increases in NVT invasive disease following routine PCV use [53–57]. The increase in NVT invasive disease after vaccine introduction is termed serotype

replacement [58]. Two mechanisms of serotypes replacement have been described: first is the increase in NVTs to partially occupy the niche evacuated by VTs after vaccine introduction [58,59]. Second is capsular switching where a *cps* locus encoding a VT capsule is replaced by a *cps* locus encoding a NVT capsule through homologous recombination [60–63]. Examples of serotype replacement include the increase of serotype 19A to become the predominant disease-causing serotype in several countries after the introduction of PCV7 [50,64–67]. However, in some countries, serotype 19A disease increased even before PCV introduction [68]. Elevated 19A carriage rates, capsular switching, and penicillin-nonsusceptibility were reported as possible factors contributing to the expansion of this serotype [69]. Serotype 35B is another example of successful post-PCV serotype replacement. Following PCV13 introduction in the USA, capsular switching from VTs 9V, 14, and 19A to NVT 35B contributed to the expansion of serotype 35B which became an important cause of IPD after PCV13 introduction [59,70,71].

1.5. Treatment of pneumococcal disease – antibiotic therapy

In South Africa, initial antibiotic therapy for pneumococcal disease includes increased doses of amoxicillin for outpatients. Hospitalised patients are empirically treated with penicillin or ampicillin, unless they are ≥65 years, or have comorbidities in which case oral amoxicillin-clavulanate or oral cephalosporins are used [37]. In the case of severe β-lactam allergy, fluoroquinolones are used as an alternative. Benzylpenicillin is recommended if the organism is susceptible, and ceftriaxone is an alternative for nonsusceptible strains and empiric treatment [72].

1.6. Antimicrobial resistance in *S. pneumoniae*

For decades, antimicrobials have been effective in the treatment of pneumococcal infections. Penicillin has been instrumental in the reduction of pneumococcal disease morbidity and mortality since the 1940s [73]. However, the widespread use and over prescription of antibiotics, especially penicillin contributed to the emergence of penicillin-resistant pneumococci in the 1960s [73–75]. In 1967, a penicillin-resistant pneumococcal strain was first described in an adult patient in Australia [75]. Subsequently, a multidrug-resistant pneumococci isolated from a child showing nonsusceptibility to three or more antibiotics was described in 1977 in South Africa [76]. Since then, the emergence of resistance to penicillin and other antibiotics has been reported globally, including countries such as France [77], Hungary [78], Iceland [79], South Africa [80], Spain [81], USA [82], and several Asian countries [83].

Commonly described antimicrobial resistance determinants for pneumococci include a combination of mutations in the transpeptidase region of three penicillin-binding protein (PBP) genes: *pbp1a, pbp2b and pbp2x* that confer resistance to β-lactam antibiotics [70]. Other resistance determinants include those for chloramphenicol inferred based on the presence of the chloramphenicol acetyltransferase gene, *cat;* erythromycin, based on the presence of erythromycin resistance methylase gene (*ermB*) or macrolide efflux pump gene (*mefA*); clindamycin, based on the presence of *ermB*; tetracycline, based on the presence of *tet*M, *tet*O or *tet*(S/M) genes; trimethoprim-sulfamethoxazole (cotrimoxazole), based on the detection of the mutation I100L in *folA* and/or indel within amino acid residue 56-67 in *folP* [70].

Over the past two decades, vaccination with PCV has effectively reduced pneumococcal disease due to both antibiotic-susceptible and antibiotic-resistant pneumococci. Several studies, including those from South Africa have shown the early benefits of reductions in antimicrobial-resistant IPD associated with VT reductions after the introduction of PCV [22,58,84–86]. However, increases in antibiotic-resistant IPD due to NVTs have been reported. For example, non-PCV13 multidrug-resistant serotype 24F emerged in France after PCV13 introduction and was expressed by a lineage that previously predominantly expressed PCV13 serotype 14 [87]. Non-PCV13 serotype 35B has long been associated with penicillin-resistance in the USA before the introduction of PCV [88], and was associated with multidrug resistance (MDR) after PCV13 introduction [84]. In 2013, three years after PCV13 introduction, serotype 35B accounted for 59.9% of MDR NVT isolates among individuals of all age groups in the USA [89]. In France, non-PCV13 serotype 15A, 35B, and 24F contributed 37.8% and 31.9% of penicillin-resistant and erythromycin-resistant isolates, respectively, with resistance to chloramphenicol and cotrimoxazole predominantly associated with serotype 12F and 24F (90,91).

A model showing that vaccination with PCV can lead to rapid increases in the frequency of pre-existing NVT resistance variants following the removal of VT competition was developed [90]. A cost analysis study showed that after PCV13 introduction in the USA (in 2012), antibiotic-resistant pneumococcal infections accounted for over 32,000 additional outpatient visits and approximately 20,000 pneumococcal pneumonia hospitalisations with incremental cost of antibiotic resistance amounting to ~$91 million in direct medical costs [91].

1.7. Molecular characterisation of *S. pneumoniae*

1.7.1. Seven-locus sequence typing

In 1998, multilocus sequence typing (MLST) was established as a standard for bacterial molecular typing [92]. MLST has been successfully used for the global surveillance of many bacterial pathogens and has been the gold standard for molecular typing for over two decades [93,94]. MLST has been traditionally used to define pneumococcal genotypes that allow the grouping of related strains into genetic clones used to describe the genetic population structure and evolutionary dynamics of *S. pneumoniae*. The allelic profile generated by sequencing the relatively conserved seven-housekeeping genes (*aroE*, *gdh*, *gki*, *recP*, *spi*, *xpt* and *ddl*) defines a sequence type (ST). Closely related STs are grouped into a clonal complex (CC) using the eBURST algorithm for clustering and identification of the founder ST in each group [95].

After PCV7 introduction, MLST was used to show that the increase in serotype 19A was due to the expansion of the pre-existing CC199 clone [96]. This penicillin intermediately-resistant clone was later outcompeted by MDR CC320/271 clone in the USA [97]. Similarly, MLST revealed that the capsular switching from VTs 9V, 14, and 19A to NVT 35B occurred within the multidrug resistant ST156 clone [59,70,71]. In South Africa, NVT 35B emerged as one of the most common causes of IPD in children <5 years after PCV13 introduction and this emergence was due to the expansion of a pre-existing penicillin-nonsusceptible clone, CC172 [98,99].

Pneumococcal disease epidemiology and the impact of PCVs thereof, have been largely explored in South Africa over the past decade. To a lesser extent, molecular epidemiology

studies have also been conducted for the pre-PCV [100–102] and PCV13 periods [102]. Predominant genotypes that were circulating pre-PCV included CC458, CC32, CC289, CC473, CC156, CC2185, CC218, CC4881, CC230, CC1016, CC2062, CC320, and CC242 for PCV13 serotypes 3, 4, 5, 6A, 6B, 6C, 7F, 9V, 14, 18C, 19A, 19F and 23F, respectively [100]. Globally disseminated lineages such as CC217 and CC230 were predominantly circulating in South Africa [100,102], and the hyper-virulent African lineage CC217 has been implicated with meningitis outbreaks in Sub-Saharan Africa [103]. For commonly circulating NVTs in South Africa such as serotype 8 and 15B/C, the corresponding genotypes were ST53 and ST7052, respectively, pre-PCV [100].

1.7.2. Global Pneumococcal Sequence Clusters (GPSC)

Due to the inclusion of only seven housekeeping genes, MLST is sometimes limited in its ability to differentiate lineages as common ancestry can be concealed by recombination events and genetic variations that may have accumulated across the genome over extended periods of time [104,105]. Therefore, MLST potentially lacks the necessary resolution to identify distinct outbreaks caused by closely related bacterial variants, or evolution within lineages [106]. The introduction of high throughput sequencing technologies has led to increasingly affordable and practical whole-genome sequencing (WGS) analyses. These technologies have enabled isolate characterisation at high resolution which facilitated the adoption of WGS analysis for routine genomic surveillance and potentiated the detection and tracing of outbreaks as they occur [106–108].

Software such as the Population Partitioning using Nucleotide K-mers (PopPUNK) have made it possible to exploit both the core and accessory genomic variations of bacterial

populations and clustering of closely related isolates [109]. The ability of PopPUNK to process 10^3-10^4 whole-genomes in a single batch, has facilitated real-time genomic surveillance of bacterial pathogens and outbreak detection in minutes. The Global Pneumococcal Sequencing (GPS) project used PopPUNK to cluster a GPS dataset of 20,027 isolates into 621 lineages referred to as Global Pneumococcal Sequence Clusters (GPSCs) [105]. These lineages were supported by the clustering comparator HierBAPS [110]. GPSCs identified shared history that was not captured by the conventional MLST CC designations. For example, clustering revealed shared descent of CC53, CC1012, CC62 and CC100 within GPSC3, which accounted for all IPD cases caused by serotype 8 pneumococci, the most common NVT in South Africa in the PCV13 era [105].

GPSC lineages were used to contextualise the distribution of serotypes, STs, and AMR determinants among pneumococcal strains which aided in assessing the impact of PCV at a global level [58]. For example, GPSC9 (ST63) switched from commonly expressing VT 14 in Cambodia, Israel, Malawi, South Africa, and the US pre-PCV to largely expressing NVT 15A in all these countries, except Malawi post-PCV [111]. In Malawi, however, GPSC9 (VT 14, ST63) was outcompeted by the expansion of GPSC9 (VT 14, ST782) post-PCV [112]. In South Africa, the historical MDR GPSC10 lineage (Denmark^{14}-32, CC230) that predominantly expressed VT 14 pre-PCV, expanded to express additional MDR NVTs 10A, serogroup 24, and 23A after PCV13 introduction [99]. A look-up table for GPSCs, CCs, and STs circulating in South Africa is included in Appendix B.

1.7.3. Bioinformatics tools for *S. pneumoniae* genomic analysis

S. pneumoniae is a naturally transformable pathogen and temporal fluctuations in the prevalence of serotypes and genotypes may occur naturally in the absence of pneumococcal conjugate vaccine pressures [113–115]. Additionally, changes in response to vaccine pressures augment the inherent variability of the pneumococcus, hence the need for genomic surveillance using tools that offer high molecular characterisation resolution. A study using WGS for genotyping examined penicillin-resistance in pneumococcus and demonstrated previously undetected genetic exchanges in the penicillin-binding genes using fastGEAR [116]. FastGEAR can accurately identify known and potential recombination hotspots, making it an appropriate tool for the detection of emerging resistance determinants. Other WGS tools such as Gubbins allow for the detection of genetic changes that result from both horizontal sequence transfer (HGT) and SNPs, which confer antibiotic-resistance among closely related pathogens [117].

There are currently several bioinformatics pipelines that are used globally to characterise the pneumococcus using WGS data. Pipelines such as those developed by Metcalf and colleagues [70], use software such as the short read typing tool SRST2 to detect STs and identify genes related to both serotype classification and antibiotic-resistance [118]. Within this pipeline, ARG-ANNOT is incorporated for the detection of novel AMR genes [119]. The pipeline was developed for real-time genomic surveillance of the pneumococcus in the Active Bacterial Core surveillance program, USA, CDC. Other WGS pipelines include the Sanger serotyping pipeline SeroBA that identifies pneumococcal serotypes from WGS raw data [120]; MLSTcheck, a fast, robust and scalable tool that outputs STs of multiple isolates in one run [121]; ARIBA, a pipeline that identifies AMR-associated genes and SNPs directly from WGS short reads [122]. Unlike the widely-used SRST2 that is limited in identifying resistance that

is conferred by the presence of a gene or a pre-defined allele of a gene, ARIBA also identifies and interprets SNP-variants that confer resistance. The global and rapid expansion of the utilisation of WGS technologies and bioinformatics tools for pathogen genomic surveillance warrants their exploitation in the South African context.

1.7.4. Tools for phylogenetics and evolutionary inference

Phylogenetics is an important approach for reconstructing the evolutionary history of an organism by inferring ancestral lineages and identifying evolutionary events such as divergence, adaptation and recombination. In the past decade, several tools have been developed to improve evolutionary inference. Gubbins [117], a maximum likelihood-based tool that accounts for the confounding effect that homologous recombination has on phylogenetic inference, detects and masks recombinant regions in bacterial whole-genome alignments for the identification of clonal relationships rather than HGT. A previous study used Gubbins to identify recombination blocks in GPSC lineages, and to create recombination-free phylogenies for the inference of evolutionary histories [105].

BactDating [123] is a Bayesian framework used to estimate ancestral dates on bacterial phylogenetic trees. The main function in the BactDating R package includes parameter estimates such as mu (μ) which represents the substitution rate (expected nucleotide changes per site per unit time), sigma (σ) which captures branch-specific mutation rate heterogeneity in the relaxed clock model (allows substitution rates to vary among branches in order to capture rate heterogeneity), and alpha (α) which corresponds to the coalescent rate. Compared to other Bayesian methods for building dated phylogenies, BactDating is faster and scalable as it assumes phylogenetic relationships have been previously established [123].

Skygrowth [124] is a non-parametric phylodynamic method used to infer past population dynamics from time-calibrated phylogenies by using a Bayesian approach to estimate the effective population size as a function of time . Similar to Skygrowth, CaveDive [125] is another method that uses Bayesian inference to reconstruct the demographic history of structured populations such as a bacterial clones using time-calibrated phylogenies to estimate clonal expansion.

1.8. Problem statement and importance of current study

S. pneumoniae remains an important cause of lower respiratory infection, bacteraemia, and meningitis, worldwide. Out of 18 pathogen categories studied in 2021 at a global scale, the pneumococcus was responsible for the highest proportions of lower respiratory infection episodes and deaths, with an estimated ~97 million episodes and over 500,000 deaths, globally, despite having 152 out of 194 World Health Organization (WHO) member states using PCV in the same year. In South Africa, PCV substantially reduced the incidence of IPD caused by serotypes targeted by the vaccine. However, increases in NVT IPD have been observed in the post-PCV era. The impact of PCVs have been delineated through epidemiological studies in this setting with the most recent findings showing the long-term effect of VT IPD reductions and NVT IPD fluctuations during the 2005–2019 period. Genomic epidemiology studies spanning this period are needed to better understand the evolution of *S. pneumoniae* following antibiotic- and vaccine-selection pressures. The pneumococcus is a naturally transformable pathogen and temporal fluctuations in the prevalence of serotypes and lineages may occur naturally in the absence of PCV pressures, but more so in response to vaccine pressures which augment the inherent variability of the pathogen. Therefore, genomic surveillance is importance to understand the impact vaccination has on the genomic population structure, evolution and adaptation of the pneumococcus. This body of work aimed to provide a detailed genomic population structure of circulating serotypes, genotypes, lineages, and antimicrobial resistance patterns of *S. pneumoniae* before and after PCV introduction in South Africa.

1.9. Study Aim and Objectives

Aim

The main aim of this thesis was to perform retrospective genomic analysis of invasive *S. pneumoniae* isolates collected through national, laboratory-based surveillance for IPD in South Africa, before and after the introduction of PCVs (2005–2020).

Objectives

1. To determine the pre- and post-PCV genomic population structure of *S. pneumoniae* by describing circulating serotypes, STs, GPSCs, and AMR determinants using whole-genome sequencing data.

2. To evaluate AMR trends among VTs and NVTs and identify lineages that express NVTs and substantially contribute to resistant invasive disease among persons of all ages after vaccine introduction.

3. To identify serotype replacement by exploring trends and genomic clustering of invasive disease-causing GPSC3 lineage that expresses the most common NVT, serotype 8 in South Africa after PCV introduction.

Chapter Two - Methods and materials

This chapter describes IPD surveillance in South Africa, study population, laboratory and genomic sequencing methods. Objective-specific methods are included in relevant chapters.

2. IPD surveillance, study population and laboratory methods

2.1. GERMS-SA surveillance programme

IPD surveillance commenced in 1999 in South Africa as part of GERMS-SA (Group for Enteric, Respiratory and Meningeal Disease Surveillance), an active, national, laboratory-based surveillance programme that facilitates the submission of invasive pneumococcal isolates and patient demographic data to the National Institute for Communicable Diseases (NICD), a division of the National Health Laboratory Service (NHLS) [52]. National surveillance was enhanced in 2003 to include additional data such as HIV serostatus and patient outcome from about 30 sentinel sites in all provinces. From 2003, surveillance reporting increased and then stabilised in 2005 [29]. Through GERMS-SA, approximately 200 laboratories across all nine South African provinces participate in IPD surveillance (Figure 2.1). An IPD case was defined as detection of *S. pneumoniae* from a normally sterile-site specimen (e.g., blood, cerebrospinal fluid [CSF], joint fluid, pleural fluid), either positive by culture, PCR, or latex agglutination plus Gram stain.

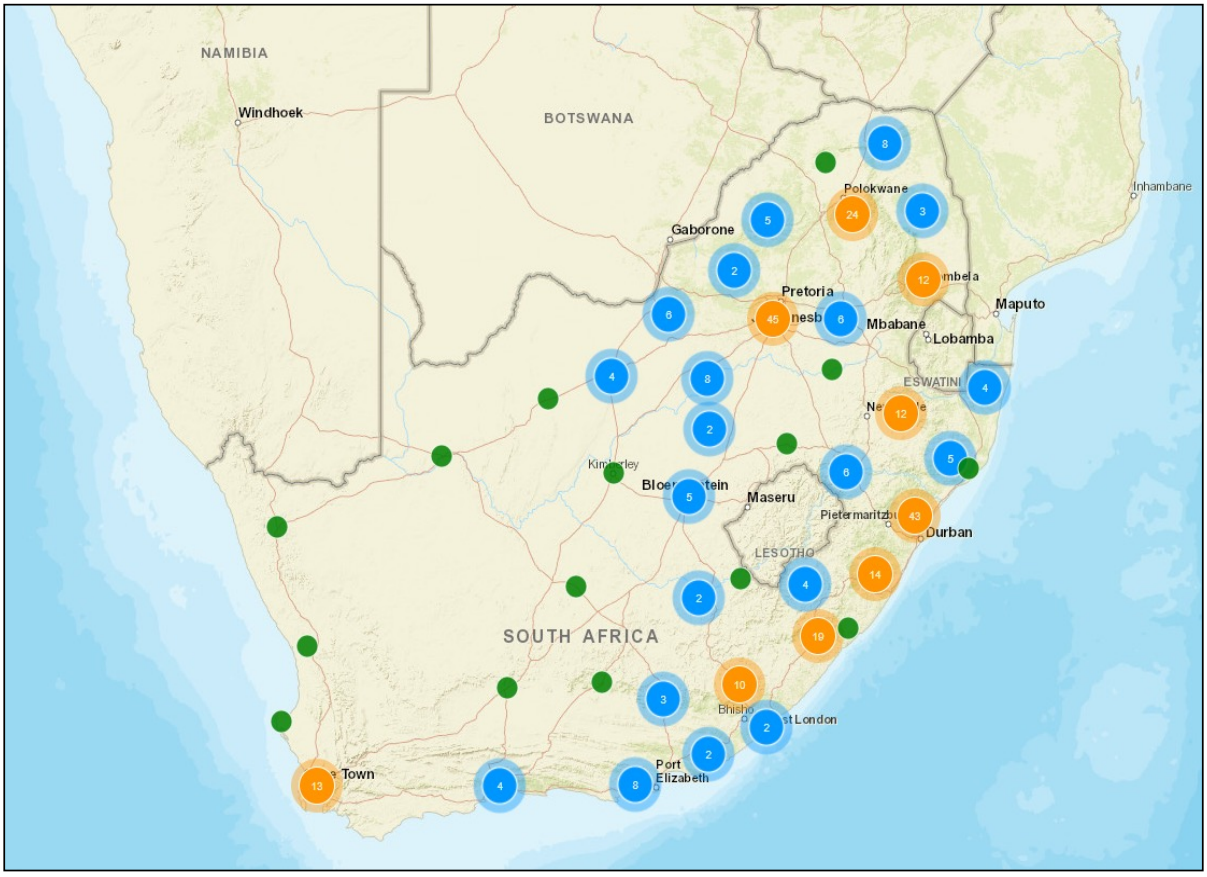


Figure 2.1: Map of South Africa showing GERMS-SA participating laboratories across all nine provinces. Green dots indicate the presence of one laboratory in the area. Blue circles are indicative of 2-9 laboratories in a region. Orange indicates ≥10 laboratories.

2.2. Isolate processing and phenotypic characterisation

Pneumococcal isolates are routinely transported from diagnostic laboratories on Dorset transport medium (Diagnostic Media Products [DMP], Johannesburg, South Africa) to the NICD reference laboratory. On arrival, *S. pneumoniae* identification is confirmed by characteristic growth when sub-cultured on 5% sheep blood agar plates (DMP, Johannesburg, South Africa) at 37ºC for 18–24 hours at 5% CO_2, and sensitivity to optochin disc (Mast Diagnostics, Virginia, USA). Pure cultures are cryopreserved in 10% skim milk (DMP, Johannesburg, South Africa) at -70ºC. Normally sterile-site specimens are confirmed using

antigen latex agglutination consistent with Gram stain or PCR [126]. IPD cases include unreported laboratory-confirmed cases identified through database audits in participating laboratories. Audit cases were grouped into a 'no isolate available' category together with cases identified on PCR only (culture-negative) or cases where the isolate was never sent from the diagnostic laboratory or lost viability during transit, laboratory processing or storage.

S. pneumoniae serotypes for viable isolates are routinely identified using serotype-specific antisera (SSI, Diagnostica, Copenhagen, Denmark) of the Quellung reaction [127]. PCR is used for *in silico* serotyping of culture-negative samples which includes clinical specimens and Dorset transport medium samples that lose viability during transportation to the NICD [128–130]. For antimicrobial susceptibility testing (AST) all isolates are initially screened for nonsusceptibility using the direct Kirby-Bauer disk diffusion method. From 2005 to 2008, AST was performed using agar dilution for penicillin and ceftriaxone, and Etest for amoxicillin, clindamycin, erythromycin, chloramphenicol, cotrimoxazole, linezolid, ofloxacin, rifampicin, tetracycline, and vancomycin (AB Biodisk, Solna, Sweden). From 2009 onwards, AST was determined by broth microdilution using commercially prepared Sensititre-SASP2 panels (Trek Diagnostics Inc., Cleveland, OH) and minimum inhibitory concentrations (MIC) were interpreted according to the Clinical and Laboratory Standards Institute (CLSI) guidelines and breakpoints [131]. Penicillin nonsusceptibility is at MIC ≥0.12 μg/mL using meningitis breakpoints. For the remaining antimicrobials, nonsusceptibility is defined at intermediately- or fully-resistant according to CLSI breakpoints. Multidrug resistance is defined as nonsusceptibility to three or more antimicrobial classes.

2.3. Study population and sampling strategy

From 2005 through 2020, 54,199 IPD cases were reported to the NICD. Of these 96% (51,917/54,199) had known age of which 33% (17,121/51,917) were cases from children <18 years and 67% (34,795/51,917) adults ≥18 years of age. Among individuals <18 years 72% (12,283/17,121) of the cases were culture-positive at the primary laboratory. Of these, 96% (11,833/12,283) had viable isolates. For persons ≥18 years, 75% (26,167/34,795) of the cases were culture-positive, of which 90% (23,586/26,167) had viable isolates. (Figure 2.2).

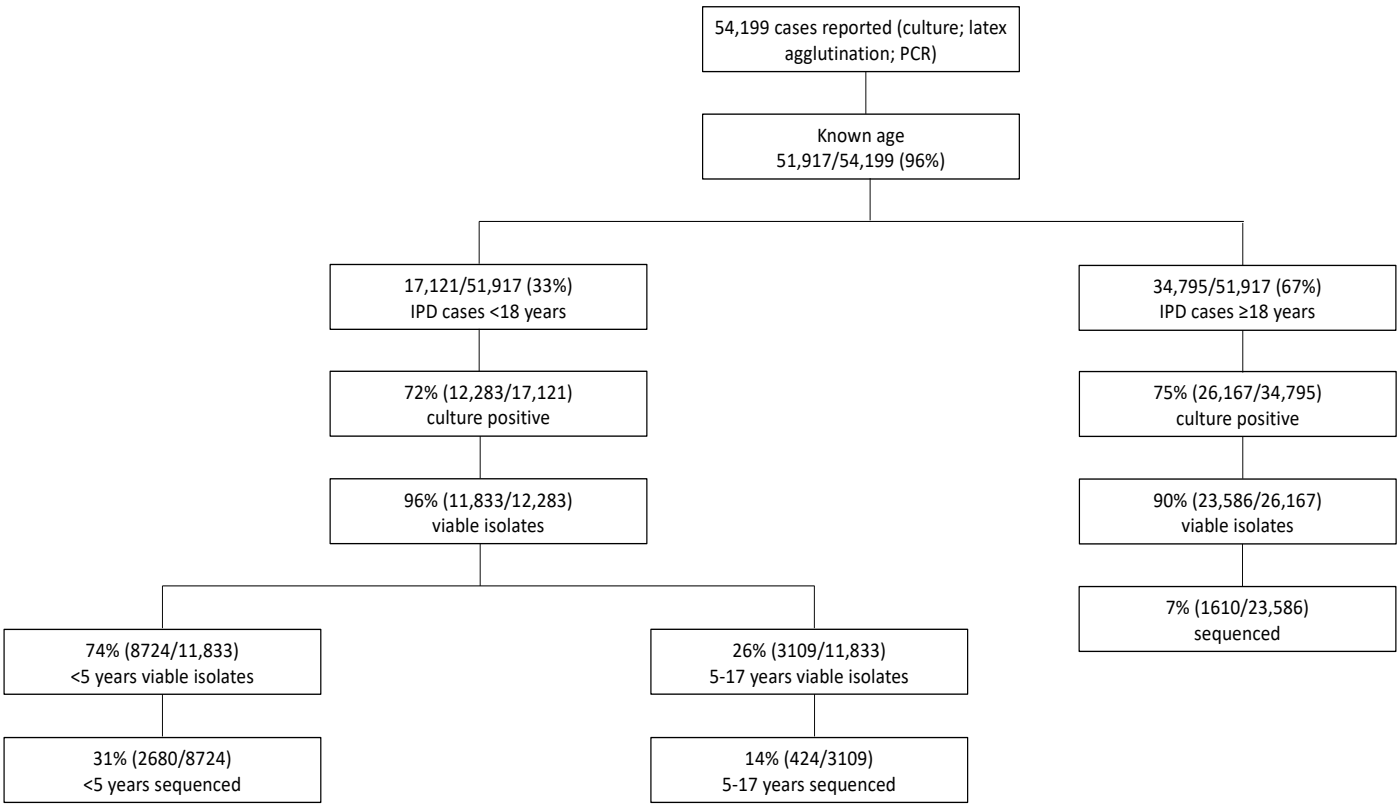


Figure 2.2: Flowchart showing total numbers of IPD cases and sequenced isolates from persons of all ages during the 2005–2020 period in South Africa.

The sampling strategy for WGS was pre-defined as part of the GPS project and was purposively biased towards the <2 year olds due to the higher disease burden in this age group that is targeted for vaccination. Viable IPD isolates were stratified into three age categories for sequencing, namely, 0–2, 3–5, and >5 years, with sample proportions of 50%, 25%, and 25% of the isolates, respectively. Approximately 300 isolates, assumed to represent all serotypes, were randomly selected per year from 2005–2020. In alignment with the sampling strategy ~150/300 (50%) isolates in the 0–2 years age category were selected and ~75/300 (25%) isolates were selected for each of the 3–5 and >5 years age categories, respectively. However, due to the substantial decrease in IPD incidence in individuals <5 years of age in the post-PCV13 period and the large number of isolates from those ≥5 years, we selected ~150 isolates from the combined 3–5 and >5 years age categories after PCV13 introduction, such that >75 isolates from individuals >5 years were selected when isolates in the 3–5 years age group were <75. The age categories used for sampling are different from those in subsequent chapters as informed by epidemiologic relevance.

2.4. DNA extraction and whole-genome sequencing

S. pneumoniae DNA extraction was performed at the NICD. DNA extraction for isolates collected from 2005–2014 was previously completed as part of the GPS project (phase I). DNA extraction for isolates collected from 2015–2020 was performed by the student (phase II). Viable cultures from cryopreservation were inoculated into Brain Heart Infusion (BHI) broth (DMP, Johannesburg, South Africa) and incubated overnight at 37°C in 5% CO_2. The BHI cultures were then checked for purity by inoculating on 5% sheep blood agar (DMP, Johannesburg, South Africa) overnight under the same conditions. Pure BHI culture cells

were sedimented by vortexing and resuspended in 10 mg/mL of lysozyme (Sigma-Aldrich, St. Louis, MO) for pre-lysis at 37°C for 1 hour. Genomic DNA was manually extracted using QIAamp® DNA Mini kit (QIAGEN, Hilden, Germany) as per manufacturer's instructions. Extracted DNA was quantified using the Qubit® 2.0 instrument (Invitrogen, Carlsbad, CA, USA) and Qubit™ dsDNA BR assay kit (Invitrogen) as per manufacturer's instructions.

4412 IPD isolates from South Africa (2005–2019) were successfully sequenced at the Wellcome Sanger Institute (Hinxton, UK) using either HiSeq or NovaSeq Illumina platforms (Illumina, San Diego, CA, USA). In addition, 302 isolates, collected in 2020, were sequenced at the NICD Sequencing Core Facility (Johannesburg, South Africa). Briefly, preparation of genomic DNA libraries for sequencing was done using the Illumina Nextera DNA Flex Library Preparation Kit and genomic sequencing was performed using the NextSeq 550 platform (Illumina, San Diego, CA, USA) under the 2 x 100-bp paired-end mode.

2.5. Genomic data processing and bioinformatics analysis

Raw sequencing short-read data were *de novo* assembled using the Velvet short-read assembler to generate multiple assemblies by varying the k-mer read length between 66% and 90% using VelvetOptimiser [132,133]. The assembly with the highest N50 was chosen. N50 was defined as the length of the shortest contig containing at least 50% of the total nucleotides in the assembly length. Assembly improvement was performed using SSPACE by scaffolding assembly contigs [134], and GapFiller was used to fill in sequence gaps [135]. Contigs that were shorter than 300–500 bases were excluded from the assembly. The quality of the assembly was assessed by aligning the reads against the final assembly using SMALT (https://github.com/rcallahan/smalt). A good quality assembly had the following quality

control thresholds: (1) average sequencing depth of >20X to support SNPs, (2) More than 60% of the reads must map to *S. pneumoniae* reference genome ATCC 700669 (NCBI accession code FM211187) to confirm pneumococcus genome data, (3) assembly length must be between 1,900,000–2,300,000 base pairs (bp), (4) the number of assembled contigs must be <500 for a genome, and (5) <15% of total SNPs must be heterozygous SNPs. A total of 4669 genomes met the required quality parameters. Good quality assemblies were annotated using PROKKA with genus-specific databases from RefSeq [136,137].

Serotypes were determined *in silico* using SeroBA, a k-mer based pipeline that identifies the *cps* locus of *S. pneumoniae* directly from raw sequencing read data [120]. STs were derived from the genome using MLSTcheck, a scalable command-line tool that can analyse multiple *de novo* assemblies by searching multiple databases in parallel [121]. Isolates sharing at least six of the seven MLST house-keeping genes were grouped into CCs and were defined as single-locus variants (SLV). GoeBURST was used to create MLST minimum-spanning trees showing allelic profile relationships [138]. PopPUNK was used to group pneumococcal genomes into GPSC lineages using core and accessory genome distances and the GPS reference database version 6 (https://www.pneumogen.net/gps/assigningGPSCs.html) [109].

Antibiotic nonsusceptibility was predicted *in silico* using the CDC StrepLab pipeline [70]. Nonsusceptibility to penicillin was inferred using a machine-learning model that evaluates the penicillin binding-protein alleles *pbp1A*, *pbp2B* and *pbp2X* to predict penicillin MICs in mg/L and interpreted according to CLSI M100-ED28:2018 guidelines and meningitis breakpoints [139,140]. Nonsusceptibility to chloramphenicol (CHL) was inferred based on the presence of the chloramphenicol acetyltransferase gene, *cat*; erythromycin (ERY), based

on the presence of erythromycin resistance methylase gene (*ermB*) or macrolide efflux pump gene (*mefA*); clindamycin (CLI), based on the presence of *ermB*; tetracycline (TET), based on the presence of *tet*M, *tet*O or *tet*(S/M) genes; cotrimoxazole (COT), based on the detection of the mutation I100L in *folA* and/or indel within amino acid residue 56-67 in *folP*. The pipeline does not detect rifampicin resistance determinants.

Chapter Three - Impact of pneumococcal conjugate vaccines on invasive pneumococcal disease-causing lineages among South African children

3. Chapter 3 reference:

Lekhuleni C, Ndlangisa K, Gladstone RA, Chochua S, Metcalf BJ, Li Y, Kleynhans J, de Gouveia L, Hazelhurst S, Ferreira AD, Skosana H, Walaza S, Quan V, Meiring S, Hawkins PA, McGee L, Bentley SD, Cohen C, Lo SW, von Gottberg A, du Plessis M. Impact of pneumococcal conjugate vaccines on invasive pneumococcal disease-causing lineages among South African children. Nature Communications. 2024 Sep 27;15(1):8401.

3.1. Introduction

As of 2023, PCVs have been introduced in 85% of WHO member states (https://www.who.int/countries), including South Africa. Substantial reductions in VT invasive disease and subsequent serotype replacement with NVTs were reported in several countries [26,47–49,141,142]. Serotype distribution continues to vary by geographic region in the post-PCV era. The GPS project (https://www.pneumogen.net/gps/) was established in 2011 and aimed to understand the impact of vaccination on pneumococcal populations. Using genome-wide variation, GPSC lineages were defined to capture shared descent signals among pneumococcal genomes to contextualise the global distribution of serotypes and AMR for the assessment of PCV impact [58,105,143]. The GPS project currently has a collection of over 20,000 genomes from 59 countries through which the standardised, high-resolution GPSC typing scheme was established. Not only is the GPSC typing scheme concordant with

traditional MLST, this approach can also reveal shared descent among STs and CCs [105]. Previous studies from South Africa showed that some pneumococcal lineages and MLST genotypes that cause IPD in our setting have rarely been reported elsewhere [98,100,101]. Such observations necessitate the need for ongoing spatiotemporal genomic surveillance and the inference of past population dynamics to reveal the demographic histories and growth patterns of important drivers of invasive disease [144]

Settings with sustained PCV immunisation programmes have benefited from reduced antibiotic-nonsusceptible IPD [145,146]. However, the circulation of lineages that predominantly expressed VTs pre-PCV and later predominantly expressed NVTs not only contribute to IPD persistence post-PCV but may also threaten the maintenance of reduced antibiotic-nonsusceptible disease given the shared genetic background of resistance traits among VT and NVT variants. Our study aimed to perform in-depth analysis of the genomic population structure of IPD isolates from children <18 years before and after the introduction of PCVs, identify predominant lineages causing disease after vaccine introduction, and describe AMR patterns in the post-PCV13 era.

3.2. Methods and materials

3.2.1. National invasive pneumococcal disease surveillance and sample selection

IPD surveillance and sampling strategy are as described in section 2.1–2.3.

3.2.2. Whole-genome sequencing and processing

WGS and genomic data processing are as described in section 2.4 and 2.5.

3.2.3. Species-wide phylogenetic visualisation

Species-wide visualisation was performed by mapping sequencing reads from each isolate against the *S. pneumoniae* reference genome ATCC 700669 (NCBI accession code FM211187) using SMALT (https://github.com/rcallahan/smalt). SNP-sites was used to reduce the pseudo-genome alignment to variant sites [147]. The resulting SNP alignment was used to construct a maximum-likelihood phylogenetic tree using the generalised time-reversible (GTR) model in FastTree2 [148]. Microreact was used for visualization [149].

3.2.4. Lineage-specific analyses

GPSC specific phylogenies were constructed by mapping Illumina reads of isolates forming a lineage against previously prepared lineage-specific references by Gladstone and colleagues using Burrow-Wheeler Alignment (BWA) v0.7.17-r1188 [105,150]. Gubbins was used to identify recombination blocks and to create recombination-free phylogenies using the GTR model in RAxML [117,151]. Microreact and Phandango were used for visualization of phylogenies and recombination, respectively [149,152].

3.2.5. Dating and phylodynamic analysis

The BactDating R package was used to determine the presence of a temporal signal of persistent vaccine-type lineages including those that were predominantly expressing NVTs in the late-PCV13 period using recombination-free phylogenies to perform root-to-tip

regression [123]. BactDating was run with three replicates and one equal tip date through 100 million Markov chain Monte Carlo (MCMC) iterations sampled every 1000 states with a burn-in of 10 million using the additive relaxed clock model with a coalescent prior for the Bayesian inference of ancestral dates [123,153]. The three replicate MCMC chains showed convergence when analysed with the Gelman diagnostic of approximately 1 for alpha, sigma and mu using the coda R package [153]. We assessed the temporal signal by using the modelcompare function of BactDating to compare the first replicate model to a model with equal isolate dates but ran under the same parameters. We then used the effectiveSize function of the coda R package to evaluate whether the effective sample size (ESS) on the first replicate model was more than 200. Time calibrated phylogenies were used as input for the skygrowth model to determine past effective population sizes of the lineages of interest using a Bayesian MCMC approach that was run for 100 000 iterations, allowing for population size estimations of approximately 50 timepoints prior to the start of our sampling date [124]. We then used the CaveDive R package to test for evidence of within lineage clonal expansion and the model was run for 100 million MCMC iterations [125].

3.2.6. *In silico* antimicrobial resistance determinant analyses

AMR determinants and in silico MICs were determined as described in section in section 2.5. Predicted MDR was defined as resistance to three or more of the following antibiotic classes: beta-lactams (penicillin), macrolides (erythromycin), lincosamides (clindamycin), chloramphenicol, tetracyclines (tetracycline), and sulfonamides (trimethoprim-sulfamethoxazole or cotrimoxazole). This study reports AMR data predicted from the genome.

3.2.7. Statistical analysis

PCV periods were defined as: pre-PCV (2005–08), PCV7 (2009–10), early-PCV13 (2011–14), and late-PCV13 (2015–20). A vaccine-type lineage was defined as having ≥50% PCV13 serotypes and a non-vaccine-type lineage as having >50% non-PCV13 serotypes in the pre-PCV period, as previously described [58]. Rare lineages were defined as GPSCs that were neither designated vaccine-type, nor non-vaccine-type and had <10 isolates. The chi-squared test was used to evaluate the frequency of vaccine and non-vaccine serotypes between the pre-PCV and late-PCV13 periods. Cramer's V was used to measure the association between PCV periods (pre-PCV, PCV7, early-PCV13, and late-PCV13) and frequency of vaccine and non-vaccine serotypes. The study population was corrected for sampling framework by year and by age group (<5 years and 5-17 years) using selected sample proportions against the GERMS-SA database. The overall incidence rate of lineages was imputed by assuming that the proportion of lineages by year and age group as determined through WGS was similar to that of the total number of IPD cases identified through surveillance for each year and age group. Poisson regression was used to calculate incidence rate ratios (IRR) at 95% confidence interval (CI) of imputed individual GPSC lineages using IPD incidence rate estimated per year for the three vaccine periods (PCV7, early-PCV13, and late-PCV13), compared to pre-PCV period. To correct for overdispersion, if observed, the robust standard error for minor violation of the Poisson regression assumption measured by goodness-of-fit at p-value ranging between 0.01 and 0.05 was reported, or the negative binomial regression measured at Poisson goodness-of-fit p value <0.01 was used. If none of the models fit, IRR was calculated using the average annual incidence for the three vaccine periods. To avoid IRR equating to zero or infinity, a constant number of one was added to all estimated IPD cases for when a GPSC was not observed in any of the three vaccine periods. Statistical significance was at <0.05 significance level using a two-sided p-value. Benjamini-Hochberg false discovery rate

of 5% was used to correct for multiple testing when number of tests exceeded ten. Lineage and serotype diversity over time was analysed using Simpson's diversity index 1-D (SDI) with values ranging from 0 to 1; higher values indicate higher diversity. We pooled both serotype and lineage diversity indices for comparing the overall diversity patterns by PCV period between and within serotype and lineage levels. All statistical analyses were performed in R version 4.2.1. R scripts for statistical analysis are available at https://doi.org/10.5281/zenodo.13284763.

3.3. Results

3.3.1. National invasive pneumococcal disease surveillance

Among the 54 199 IPD episodes reported during the 16-year study period (2005–20), age was captured for 96% (51 917/54 199) of which 33% (17 121/51 917) were among children <18 years. Of these, 12 283 (72%) cultures were submitted to the National Institute for Communicable Diseases (NICD); 96% (11 833/12 283) had viable isolates available for sequencing. A total of 3141 (27%) viable isolates were sequenced.

Among the 3141 genomes, we determined concordance between the routine phenotypic serotyping results and *in silico* serotypes. Isolates reacting in the pool G antisera serogroup (n=32) that could not be resolved to the serotype level phenotypically were distinguished *in silico* to serotype 29 (n=1), 34 (n=4), 35A (n=1), 35B (n=19), 35D (n=3), and 35F (n=4). Serogroup 16, 22, and 25 (n=5) were resolved to serotype 16F (n=3), 22F (n=1), and 25F (n=1), respectively. Among 64 of 3141 (2%) isolates that had discordant serotyping results, 27 isolates were viable and upon repeat testing of both phenotypic and *in silico* serotyping, all were concordant. The remaining 37 that could not be repeated due to loss of viability were

excluded. Concordance of 99% (3071/3104) was achieved between phenotypic and *in silico* serotypes; 1% (33/3104) was concordant to the serogroup level. Finally, a total of 3104 isolates were included in the current study. Isolate characteristics are shown in Appendix C.

3.3.2. Serotype distribution

Among the 3104 genomes, we identified 57 serotypes and one unencapsulated isolate. Overall, the ten most predominant serotypes in our selection, in order of prevalence, were serotype 19F (244/3104, 8%), 6A (241/3104, 8%), 8 (231/3104, 7%), 14 (231/3104, 7%), 1 (228/3104, 7%), 23F (224/3104, 7%), 19A (217/3104, 7%), 6B (181/3104, 6%), 15B/C (127/3104, 4%), and 12F (117/3104, 4%). The serotype distribution among sequenced isolates was consistent with the serotype distribution among all 11 789 IPD cases with known phenotypic serotype data from our IPD surveillance during the same period (Appendix D and E).

Before the introduction of PCVs in South Africa, IPD in children was largely caused by PCV7 serotypes 14 (122/934, 13%), 23F (112/934, 12%), 6B (95/934, 10%), and PCV13 unique serotypes 6A (108/934, 12%) and 1 (105/934, 11%). Similar serotypes predominated after the introduction of PCV7. Changes in serotype predominance were observed during the early-PCV13 period where PCV13 serotypes 19A (81/891, 9%), 1 (67/891, 8%), 23F (58/891, 7%), 6A (57/891, 6%), and NVT 8 (58/891, 7%) became the five most common serotypes causing disease. In the late-PCV13 period disease-causing serotypes were largely NVTs including serotype 8 (154/855, 18%), 12F (50/855, 6%), 15B/C (46/855, 5%), 16F (40/855, 5%), and PCV7 serotype 19F (66/855, 8%). Similar serotypes circulated among the <5 and 5-17 year olds before and after PCV introduction (Appendix F). Serotype 8 was the leading cause of IPD in children <5 years in the late-PCV13 period accounting for 20% (146

of 739) of invasive disease whilst serotype 19F was the most common cause of IPD in the 5-17 year olds with a frequency of 14% (14 of 102) in the same period. Overall, we observed a strong association (Cramer's V=0.5126) between PCV periods (pre-PCV, PCV7, early-PCV13, and late-PCV13) and the frequency of observing vaccine and non-vaccine serotypes.

3.3.3. Genomic population structure and lineage-specific incidence changes

Of the 3104 isolates in our study population, 3098 (99.8) were classified into 152 existing GPSC lineages and 3032 (98%) into 629 STs. We identified six new GPSCs all of which occurred in the late-PCV13 period. Among the 152 GPSCs, 22% (n=33), 22% (n=33), 23% (n=35), and 34% (n=51) expressed only vaccine serotypes, only non-vaccine serotypes, both vaccine and non-vaccine serotypes, and rare lineages, respectively. Lineages expressing both VTs and NVTs accounted for 55% (1719/3104) of the isolates in our collection. The ten most common lineages in this study accounted for 50% (1561/3104) of the isolates (Figure 3.1). The most commonly expressed serotypes among the ten predominant lineages were: GPSC1 (serotype 19F, n=100/104 [96%]), GPSC2 (serotype 1, 226/226 [100%]), GPSC3 (serotype 8, 220/255 [86%]), GPSC5 (serotype 35B, 60/142 [42%]), GPSC10 (serotype 14, 120/155 [77%]), GPSC13 (serotype 6A, 118/126 [94%]), GPSC14 (serotype 23F, 153/167 [92%]), GPSC17 (serotype 19A, 199/202 [99%]), GPSC21 (serotype 19F, 90/91 [99%]), and GPSC41 (serotype 6A, 88/93 [95%]) and were those largely included in the PCV13 formulation. With reference to the GPS database, GPSC17, GPSC21, and GPSC41 were more prevalent among African isolates at frequencies of 92% (349/380), 100% (275/275), and 100% (129/129), respectively.

Among children <5 years, the average incidence of vaccine-type lineages decreased significantly between the pre-PCV and PCV7 periods for GPSC2 (IRR 0.7, 95% CI 0.9-0.5,

p=0.04), GPSC10 (0.6, 0.9-0.5, p=0.03), GPSC11 (0.3, 0.7-0.1, p=0.03), GPSC14 (0.7, 0.9-0.5, p=0.03), GPSC52 (0.3, 0.8-0.1, p=0.04), and GPSC77 (0.2, 0.4-0.09, p<0.001, Figure 3.2a and Appendix G). Similar to the <5 year olds, the incidence of GPSC9 (0.05, 0.4-0.007, p<0.001) and GPSC13 (0.2, 0.4-0.08, p<0.001) declined significantly in the 5-17-year age group (Figure 3.2b and Appendix H). Comparing the pre-PCV and late-PCV13 periods, both age groups showed similar incidence reduction patterns for vaccine-type lineages (Figure 3.2a and 3.2b; Appendix I and J). For non-vaccine-type lineages, reductions in invasive disease caused by GPSC3 (0.4, 0.8-0.3, p=0.03) and GPSC48 (0.3, 0.6-0.1, p=0.03) were observed when comparing the pre-PCV and PCV7 periods among the <5-year age group. In contrast, significant increases in IPD caused by GPSC3 (1.8, 1.2-2.7, p=0.008) were observed in the same age group when comparing the pre-PCV and late-PCV13 periods (Figure 3.2a).

Before PCV introduction GPSC5, GPSC9, GPSC10, and GPSC11 predominantly expressed PCV13 serotypes. After PCV introduction, especially in the early- and late-PCV13 periods, these lineages largely expressed non-PCV13 serotypes (Figure 3.2). GPSC5 (CC172) largely expressed VT 23F (n=29/33, 88%) with no observation of NVT 35B in the pre-PCV period in this collection. In the late-PCV13 period, replacement serotype 35B (CC172) was the commonly expressed serotype within this lineage (34/49, 69%) with no observation of serotype 23F. Similarly, both GPSC9 (CC63) and GPSC10 (CC230) expressed VT 14 (27/27, 100% and 59/60, 98%) in the pre-PCV and PCV7 periods (9/9, 100%; 26/27, 96%), respectively. However, in the late-PCV13 period, GPSC9 largely expressed NVT 15A (13/19, 68%); GPSC10 expressed a combined proportion of 49% (20/41) of NVT 10A (11/41, 27%) and serogroup 24 (9/41, 22%), in addition to VT 14 (12/41, 29%).

3.3.4. Phylodynamic analysis

We observed significant temporal signals ($p<0.001$) using BactDating for the persistent vaccine-type lineages of interest that continued to circulate in the late-PCV13 period, allowing for coalescent analysis. Effective sample sizes (ESS) of greater than 200 were also obtained for these lineages. Past population dynamics for GPSC5 (CC172, serotype 35B) and GPSC17 (CC2062, serotype 19A) had comparative temporal trends with steady increases in the effective population size in the years after 1990 to the end of the study period in 2020 (Figure 3.3a and 3.3d). This observation was supported by clonal expansion findings where the time-calibrated phylogeny for GPSC5 showed the highest probability for clonal expansion in 1992 (credible interval: 1981-1998). Similarly, GPSC17 had several clonal expansion events with the earliest expansion occurring in 1953 (1936-1965) followed by several others that coincided with the increasing effective population size in 1989 (1982-1994) and 1991 (1984-1996). Although the effective population sizes for GPSC9 (VT 14, and NVT 15A) and GPSC10 (VT 14, NVT serogroup 24) showed similar increases around the 1990s, a decreasing trend in the effective population size for GPSC9 was observed before the introduction of PCV7 followed by a slight increase after PCV13 introduction (Figure 3.3b). For GPSC10, however, this decrease coincided with the introduction of PCV7 in 2009 (Figure 3.3c). Clonal expansion analysis for GPSC9 and GPSC10 revealed that the VT and NVT variants in these lineages diverged before PCV introduction with NVT 15A prominently expanding in 2005 (2001-2009), shortly before PCV7 introduction (Figure 3.4a and 3.4b).

3.3.5. Simpson's diversity index trends

We further explored serotype and lineage diversity indices across the dataset which were generally high across the study period, ranging between 0.903 to 0.993 and 0.917 to 0.971,

respectively (Figures. 3.5a and 3.5c). Compared to the pre-PCV and PCV7 years, significant increases in serotype diversity were observed in the early-PCV13 years with the highest increase occurring in the PCV13 introduction year, 2011 (Simpson's diversity index [SDI] 0.993, 95% CI 0.991-0.995). Declining serotype diversity trends were evident when comparing 2011 through to the late-PCV13 years, with diversity approaching pre-PCV estimates (Figure 3.5a). However, significant increases in lineage diversity were observed between 2011 and 2013 (0.948, 0.937-0.960 vs 0.971, 0.963-0.979, Figure 3.5c). Pooled lineage diversity trends reflected those of pooled serotype diversity trends both with significant differences when comparing the PCV7 and early-PCV13 periods (pooled lineage SDI 0.954, 95% CI 0.948-0.961 vs 0.965, 0.962-0.969 and pooled serotype SDI 0.915, 95% CI 0.907-0.924 vs 0.953, 0.950-0.957). Similarly, significant differences were also observed when comparing the early- and late-PCV13 periods (0.965, 0.962-0.969 vs 0.946, 0.937-0.956 and 0.953, 0.950-0.957 vs 0.938, 0.930-0.947, Figure. 3.5b and 3.5d).

3.3.6. Antimicrobial resistance

AMR concordances comparing phenotypic antimicrobial susceptibility testing (AST) and *in silico* AMR predictions are shown in Appendix K. Among the 3104 IPD isolates, 2058 (66%) had at least one resistance determinant for the antibiotics investigated (Figure 3.1). Overall, reductions in the prevalence of predicted antimicrobial resistance among all classes of antibiotics, except chloramphenicol, were observed when comparing the pre-PCV and late-PCV13 periods (Table 3.1). In contrast, predicted antimicrobial resistance to penicillin, erythromycin and MDR increased significantly in invasive disease caused by NVTs (Table 3.1). Among the 1249 NVTs, 232 (19%) had PBP gene combinations that conferred penicillin resistance. Of these, NVT 35B (GPSC5, n=60), 15B/C (GPSC48, n=46), 15A (GPSC9/GPSC168, n=40), 10A (GPSC10, n=12), 23B (GPSC102, n=12) and serogroup 24

(GPSC10, n=9) accounted for 77% (179/232) of the isolates and these serotypes were among the predominant NVTs circulating in the late-PCV13 period. Erythromycin resistance was identified in 50 (4%) of 1249 NVT isolates with *ermB*-mediated resistance occurring in 35 (70%) of 50 cases compared to *mefA*-mediated resistance. Of the 120 MDR NVT isolates, 73 (61%) were from the late-PCV13 period (Table 3.1). The three most common lineages driving MDR increase were GPSC9 (NVT 15A; 13/73, 18%), GPSC10 (NVT 10A; 11/73, 15% and NVT serogroup 24; 9/73, 12%), and GPSC26 (NVT 12F, 30/73, 41%). GPSC9 and GPSC10 were both resistant to penicillin, tetracycline, and cotrimoxazole, with GPSC10 additionally resistant to erythromycin. All of serogroup 24 MDR isolates were from the late-PCV13 period in this study.

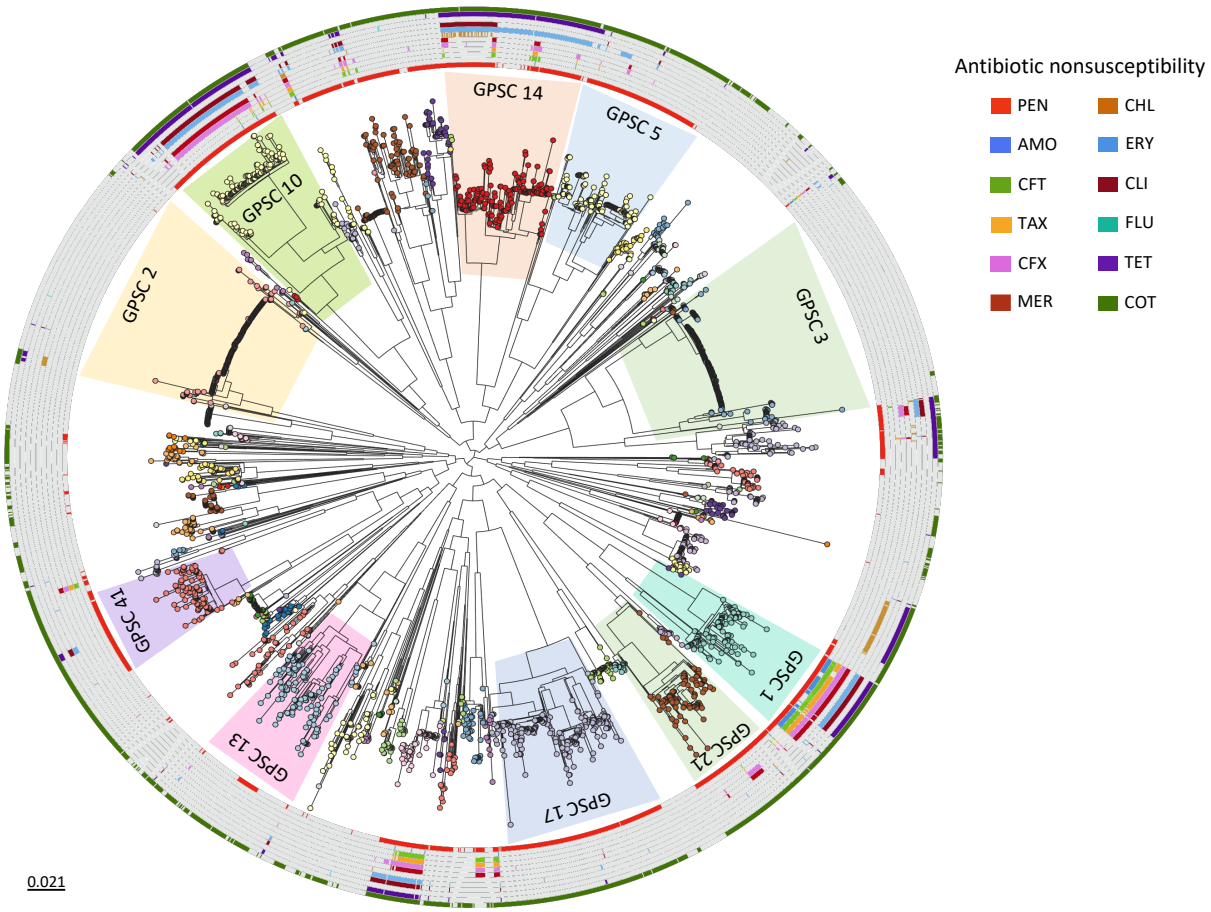
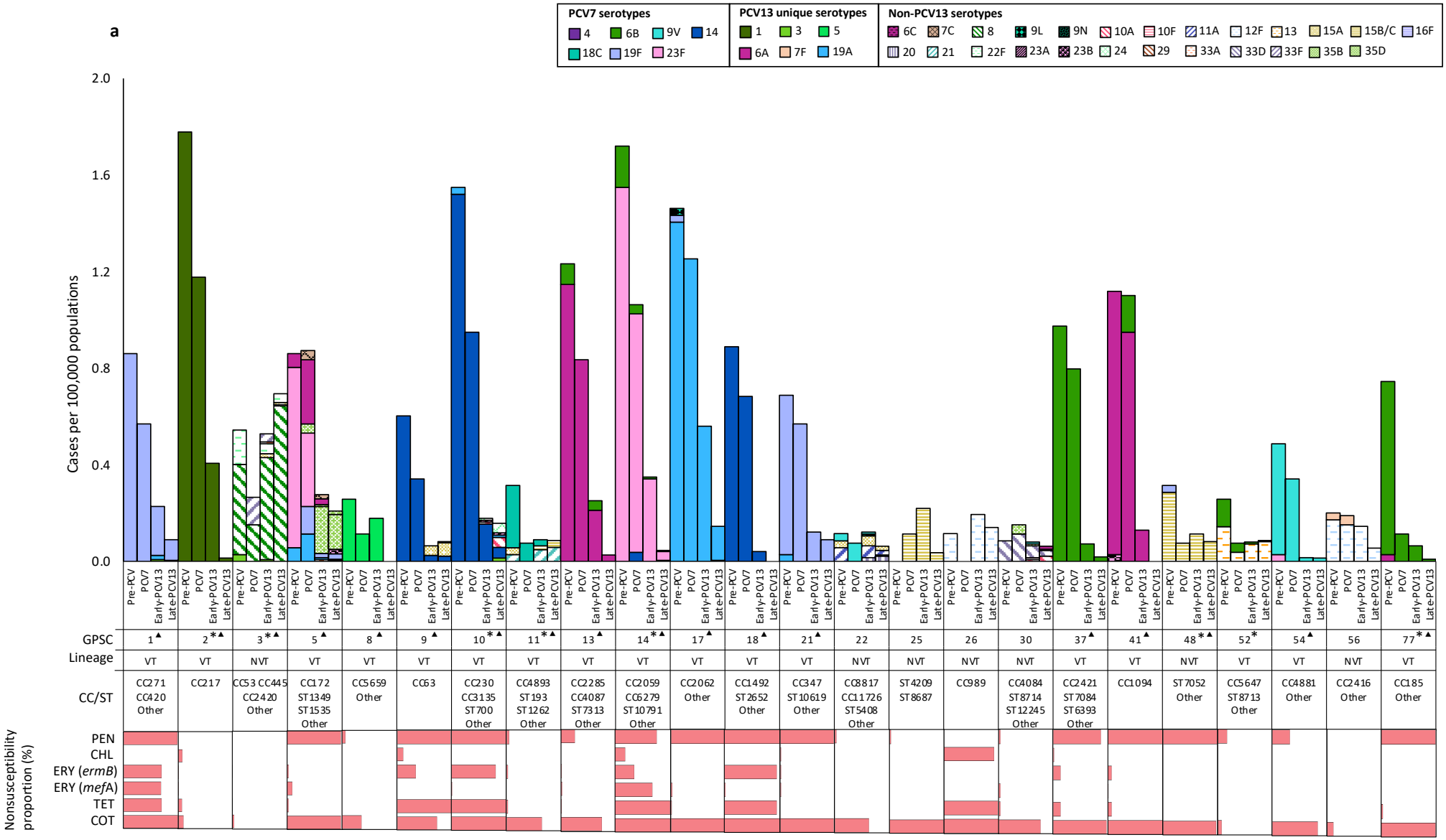


Figure 3.1: Single nucleotide polymorphism (SNP) maximum-likelihood phylogeny (N=3104) of invasive pneumococcal disease isolates from South African children <18 years. The SNP tree is midpoint rooted. The 10 most predominant GPSC taxa are highlighted and listed in descending order: GPSC3 (n=255/3104 [8%], GPSC2 (n=226/3104 [7%], GPSC17 (n=202/3104 [7%], GPSC14 (n=167/3104 [5%], GPSC10 (n=155/3104 [5%], GPSC5 (n=142/3104 [5%], GPSC13 (n=126/3104 [4%], GPSC1 (n=104/3104 [3%], GPSC41 (n=93/3104 [3%], and GPSC21 (n=91/ 3104 [3%]. Antibiotic nonsceptibility by GPSC lineage is represented by coloured circular strips: PEN=penicillin; AMO=amoxicillin; CFT=ceftriaxone; TAX=cefotaxime; CFX=cefuroxime; MER=meropenem; CHL=chloramphenicol; ERY=erythromycin; CLI=clindamycin; FLU=fluoroquinolones; TET=tetracycline; COT=cotrimoxazole. GPSC=Global Pneumococcal Sequence Clusters.


a
PCV7 serotypes
4
6B
9V
14
18C
19F
23F
PCV13 unique serotypes
1
3
5
6A
7F
19A
Non-PCV13 serotypes
6C
7C
8
9L
9N
10A
10F
11A
12F
13
15A
15B/C
16F
20
21
22F
23A
23B
24
29
33A
33D
33F
35B
35D
Cases per 100,000 populations
2.0
1.6
1.2
0.8
0.4
0.0
Pre-PCV
PCV7
Early-PCV13
Late-PCV13
GPSC
1▲
2*▲
3*▲
5▲
8▲
9▲
10*▲
11*▲
13▲
14*▲
17▲
18▲
21▲
22
25
26
30
37▲
41▲
48*▲
52*
54▲
56
77*▲
Lineage
VT
NVT
CC/ST
CC271 CC420 Other
CC217
CC53 CC445 CC2420 Other
CC172 ST1349 ST1535 Other
CC5659 Other
CC63
CC230 CC3135 ST700 Other
CC4893 ST193 ST1262 Other
CC2285 CC4087 ST7313 Other
CC2059 CC6279 ST10791 Other
CC2062 Other
CC1492 ST2652 Other
CC347 ST10619 Other
CC3817 CC11726 ST5408 Other
ST4209 ST8687
CC989
CC4084 ST8714 ST12245 Other
CC2421 ST7084 ST6393 Other
CC1094
ST7052 Other
CC5647 ST8713 Other
CC4881 Other
CC2416 Other
CC185 Other
Nonsusceptibility proportion (%)
PEN
CHL
ERY (ermB)
ERY (mefA)
TET
COT

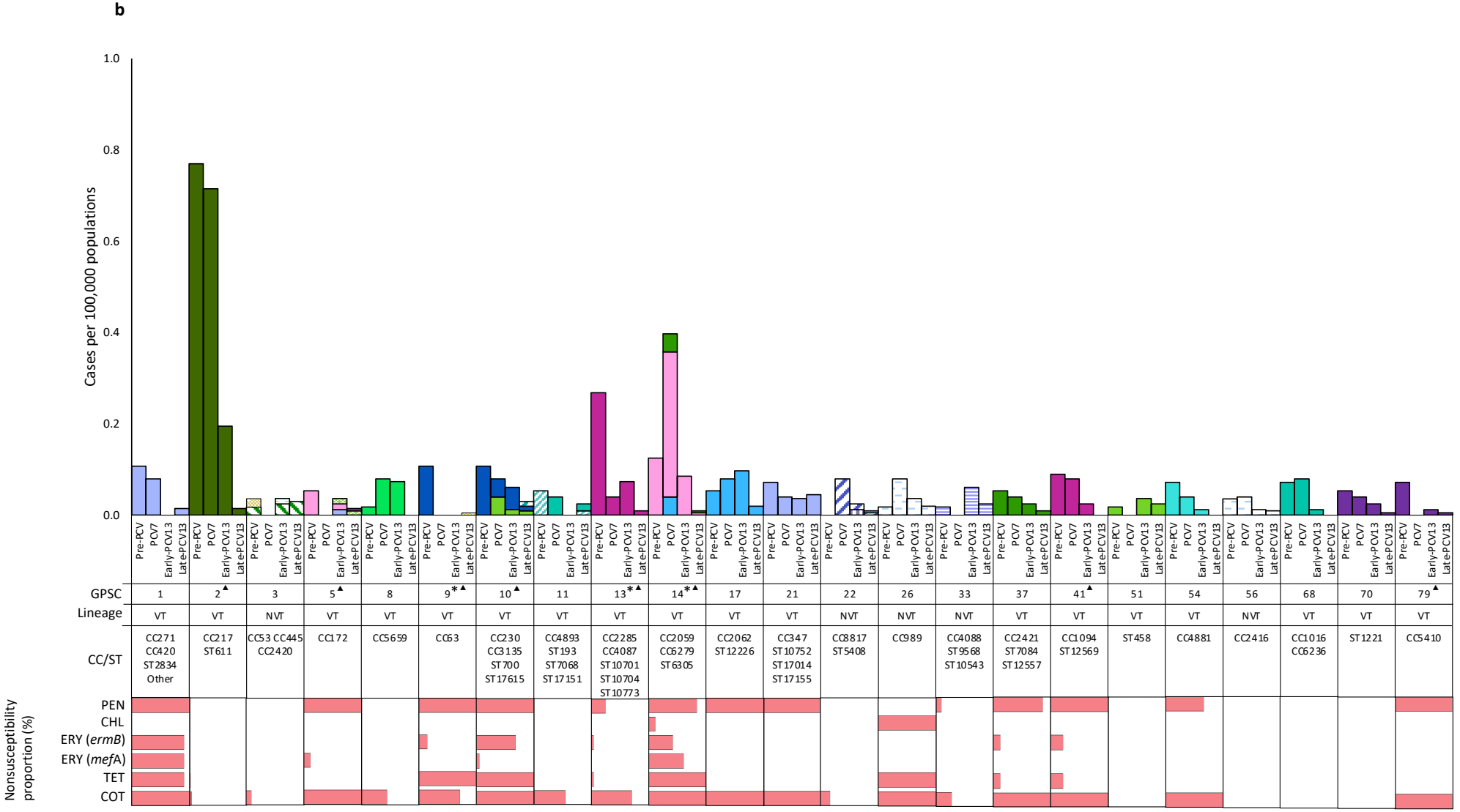

b
Cases per 100,000 populations
1.0
0.8
0.6
0.4
0.2
0.0
Pre-PCV
PCV7
Early-PCV13
Late PCV13
GPSC
1
2▲
3
5▲
8
9*▲
10▲
11
13*▲
14*▲
17
21
22
26
33
37
41▲
51
54
56
68
70
79▲
Lineage
VT
NVT
CC/ST
CC271 CC420 ST2834 Other
CC217 ST611
CC53 CC445 CC2420
CC172
CC5659
CC63
CC230 CC3135 ST700 ST17615
CC4893 ST193 ST7068 ST17151
CC2285 CC4087 ST10701 ST10704 ST10773
CC2059 CC6279 ST6305
CC2062 ST12226
CC347 ST10752 ST17014 ST17155
CC8817 ST5408
CC989
CC4088 ST9568 ST10543
CC2421 ST7084 ST12557
CC1094 ST12569
ST458
CC4881
CC2416
CC1016 CC6236
ST1221
CC5410
Nonsusceptibility proportion (%)
PEN
CHL
ERY (ermB)
ERY (mefA)
TET
COT

Figure 3.2: Global Pneumococcal Sequence Cluster (GPSC) lineages among invasive disease isolates from (a) children <5 years (n=2680) and (b) 5-17 years (n=424) across vaccine periods in South Africa (N=3104). Invasive pneumococcal disease incidence per 100 000 population is plotted by GPSC lineages stratified into four vaccine periods (pre-PCV [2005-2008], PCV7 [2009-2010], early-PCV13 [2011-2014], and late-PCV13 [2015-2020]). Serotypes included in the 13-valent pneumococcal conjugate vaccine are shown by solid colour fills; non-vaccine serotypes by hatched pattern colour fills. Lineages with <40 isolates are not shown. A vaccine-type (VT) lineage was defined as having ≥50% vaccine serotype(s) in the pre-PCV period; non-vaccine-type (NVT) lineage as having >50% non-vaccine serotype(s) in the pre-PCV period. Lineage specific clonal complexes (CC) and sequence types (ST) are shown underneath the graph. Either Poisson regression or negative binomial regression was used to calculate the incidence rate ratios (IRR) of GPSCs by vaccine period. If neither model fit, the IRR was calculated using the average annual incidence. Significant changes in the average incidence between the pre-PCV and PCV7 or late-PCV13 periods are represented by an asterisk or triangle, respectively. Significance was determined at <0.05 using a two-sided p-value where applicable. Multiple testing (>10 tests) was corrected using the Benjamini-Hochberg false discovery rate of 5%. Antibiotic nonsusceptibility proportions by lineage are shown in red horizontal bars. Penicillin (PEN) nonsusceptibility was predicted based on the penicillin-binding protein types: *pbp*1a, *pbp*2x, *pbp*2b; chloramphenicol (CHL), based on the presence of chloramphenicol acetyltransferase gene, *cat*; erythromycin (ERY), based on the presence of erythromycin resistance methylase gene *ermB* or macrolide efflux pump gene *mefA*; tetracycline (TET), by the presence of *tet*M or *tet*S/M gene with no promoter region interruptions; cotrimoxazole (COT), by the presence of mutation I100L in *folA* and/or indel within amino acid residue 56-67 in *folP*.

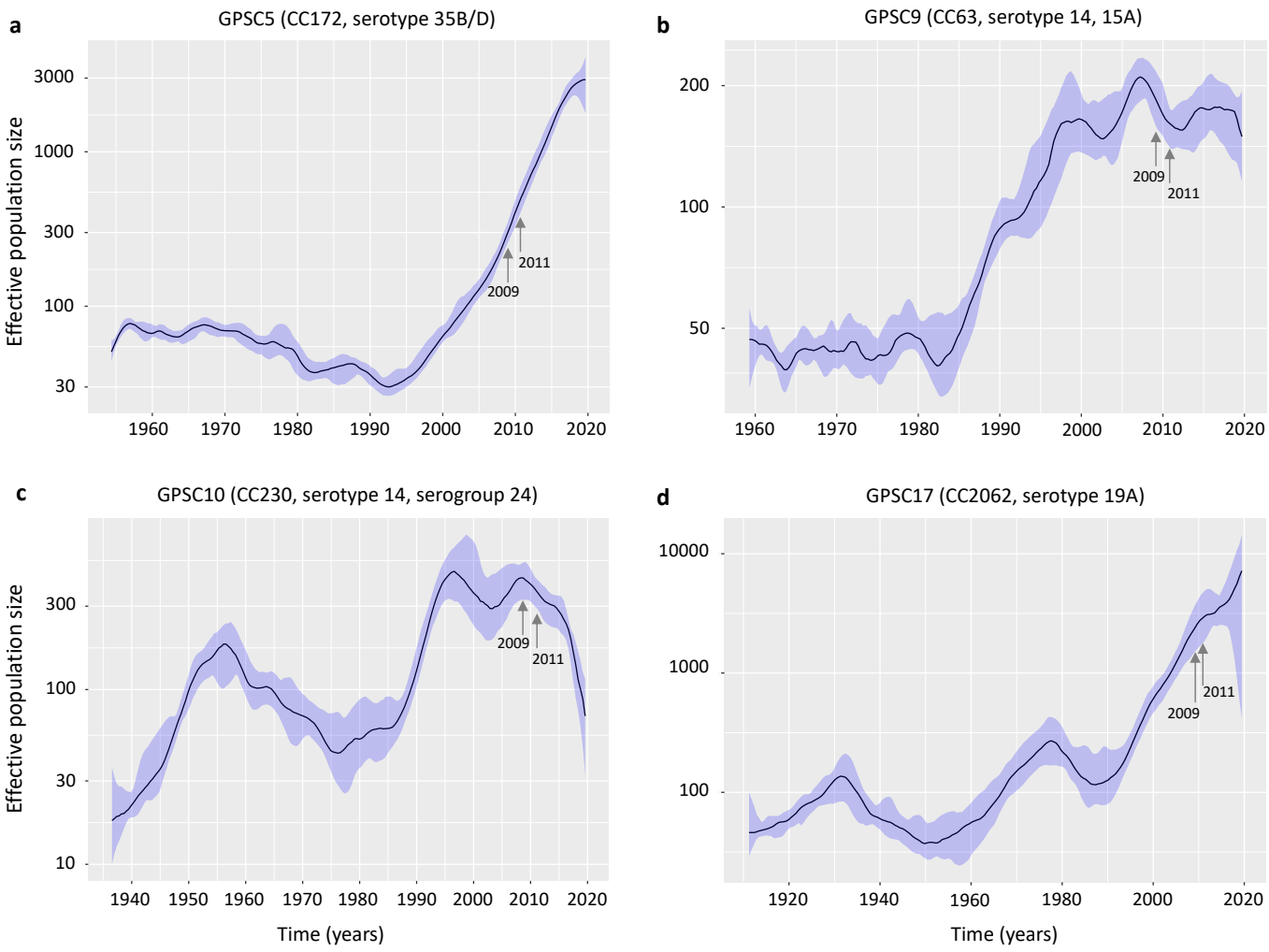


Figure 3.3: Temporal trend analysis of important vaccine-type lineages causing persistent invasive disease in the late-PCV13 period inferred using skygrowth. a-d Past population size dynamics as a function of time inclusive of the pneumococcal conjugate vaccine (PCV) introduction years shown by the grey arrows (PCV7 in 2009 and PCV13 in 2011) in South Africa. Time-calibrated phylogenies were used for inference with root dates as start time. The clonal complexes (CC) and serotypes expressed by each lineage are shown in each panel title. The solid line represents the posterior median population size. The shaded purple area represents the 95% credible intervals. The effective population size on the y-axis is expressed in log scale. GPSC = Global Pneumococcal Sequence Cluster. *Only serotypes that had 5 or more genomes are included in the a-d panel titles.

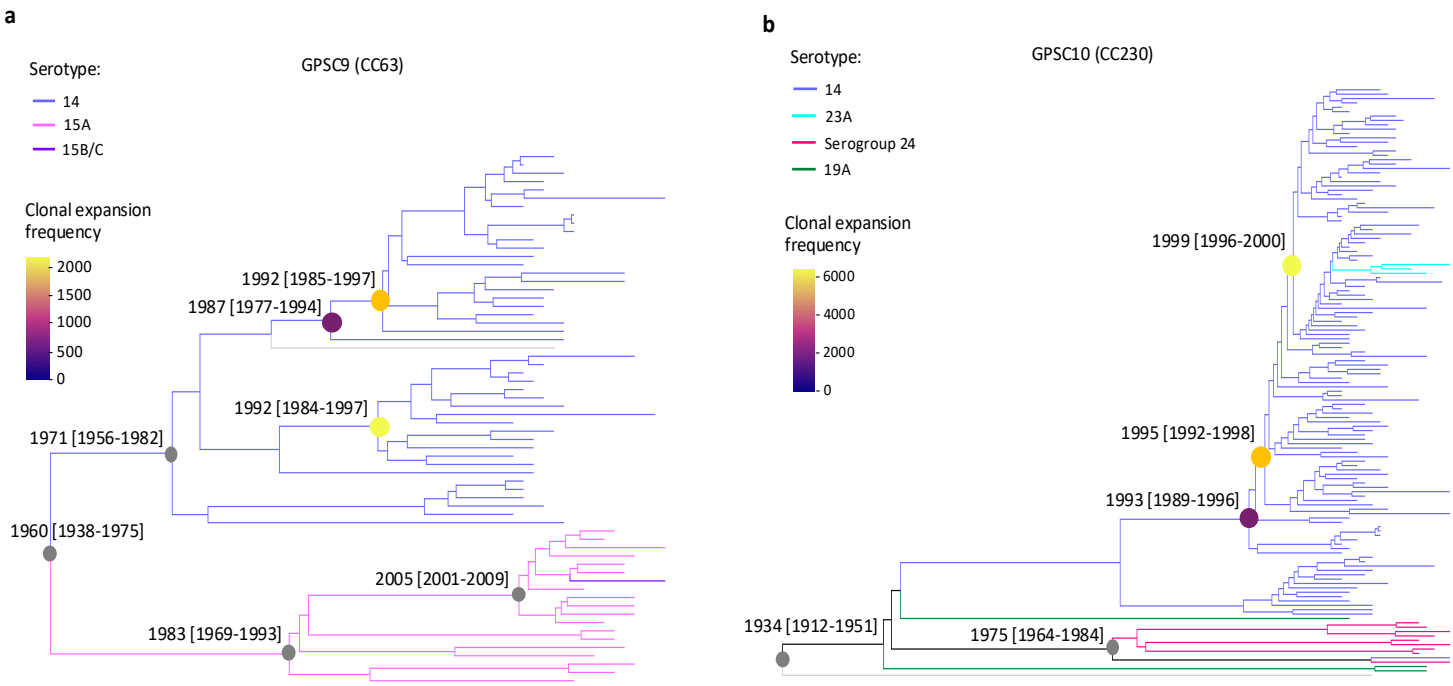


Figure 3.4: Clonal expansion dynamics of GPSC9 (a) and GPSC10 (b) inferred using CaveDive. Time-resolved phylogenies showing within lineage clonal expansions with clades coloured by serotype. GPSC references are coloured in grey. The parental nodes with high probabilities for clonal expansions are highlighted by the clonal expansion frequency colour scale and the estimated expansion year with 95% credible intervals is included. Other parental nodes are highlighted in grey including the root date. GPSC = Global Pneumococcal Sequence Cluster. CC = clonal complex.

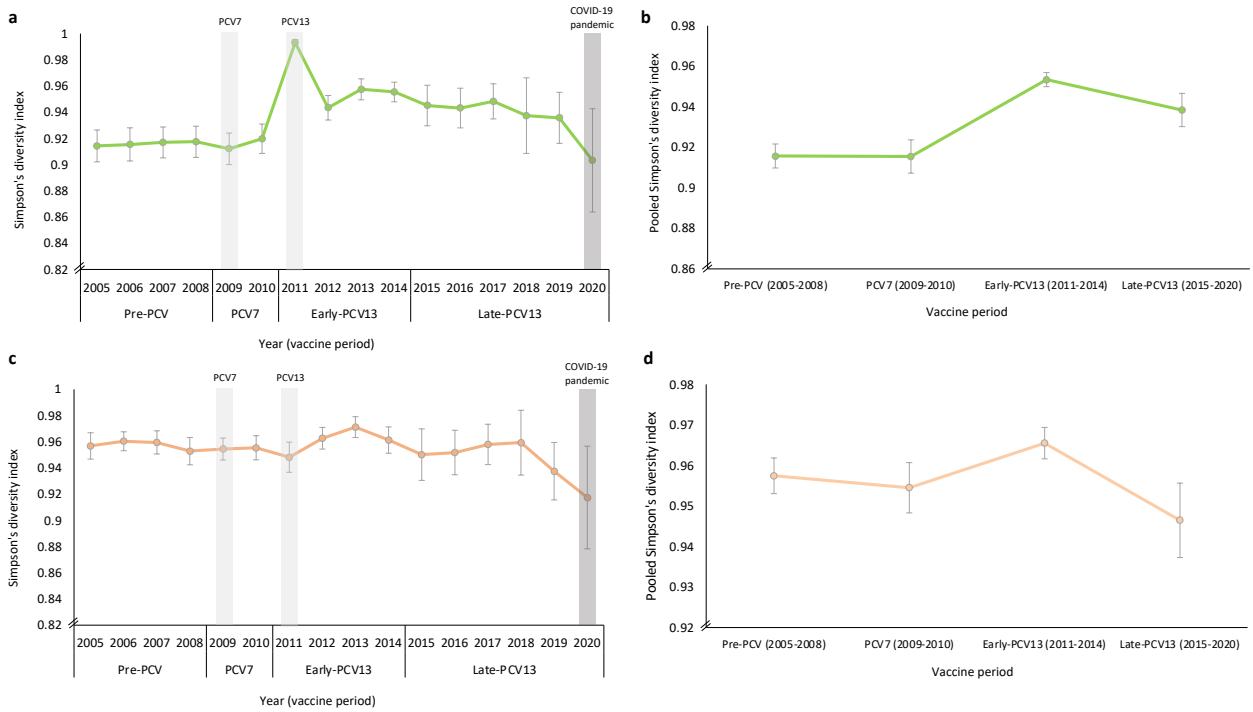


Figure 3.5: Simpson's diversity index (SDI) estimates for invasive disease serotypes and lineages. a Serotype diversity trends by year showing fluctuations before and after the introduction of seven- and 13-valent pneumococcal conjugate vaccines (PCV7 and PCV13) in 2009 and 2011, respectively (highlighted in light grey); the COVID-19 pandemic year (2020) is highlighted in dark grey. b Pooled serotype SDI estimates by vaccine period showing statistically significant differences between the PCV7 and early-PCV13 periods; early-PCV13 and late-PCV13 periods. c Lineage diversity trends by year before and after the introduction of PCV7 and PCV13 in 2009 and 2011, respectively (highlighted in light grey); the COVID-19 pandemic year (2020) is highlighted in dark grey. d Pooled lineage SDI estimates by vaccine period showing statistically significant differences between the PCV7 and early-PCV13 period; early-PCV13 and late-PCV13 periods. Estimates were derived at 95% confidence interval among invasive disease isolates (N=3104) from children <18 years in South Africa during the pre-PCV (n=924), PCV7 (n=457), early-PCV13 (n=882), and late-PCV13 (n=841) periods. SDI estimates represent the measure of diversity with standard error included at each data point.

Table 3.1: Changes in the prevalence of antibiotic nonsusceptibility among South African children, comparing the pre-PCV and late-PCV13 periods.

Antibiotic Resistance mechanism	Pre-PCV* (2005-2008)	PCV7 (2009-2010)	Early-PCV13 (2011-2014)	Late-PCV13* (2015-2020)	Adjusted p-value linear regression**
Overall (n)	924	457	882	841	
Penicillin	507 (55%)	266 (58%)	345 (39%)	293 (35%)	<0.0001
pbp:1a-2b-2x†	507 (55%)	266 (58%)	345 (39%)	293 (35%)	
Erythromycin	203 (22%)	112 (25%)	118 (13%)	86 (10%)	<0.0001
ermB or *ermB/mefA*	151 (16%)	81 (18%)	75 (9%)	66 (8%)	
mefA	52 (6%)	31 (7%)	43 (5%)	20 (2%)	
Clindamycin	151 (16%)	81 (18%)	75 (9%)	66 (8%)	<0.0001
ermB or *ermB/mefA*	151 (16%)	81 (18%)	75 (9%)	66 (8%)	
Chloramphenicol	29 (3%)	14 (3%)	46 (5%)	35 (4%)	0.27
cat	29 (3%)	14 (3%)	46 (5%)	35 (4%)	
Tetracycline	246 (27%)	119 (26%)	159 (18%)	137 (16%)	<0.0001
*tet*M	200 (22%)	102 (22%)	139 (16%)	131 (16%)	
*tet*S/M	46 (5%)	17 (4%)	20 (2%)	6 (1%)	
Cotrimoxazole	637 (69%)	338 (74%)	537 (61%)	479 (57%)	<0.0001
*folA*_I100L	17 (2%)	6 (1%)	7 (1%)	4 (0%)	
*folP*_aa_insert_57-70	170 (18%)	73 (16%)	162 (18%)	152 (18%)	
*folA*_I100L, *folP*_aa_insert_57-70	450 (49%)	259 (57%)	368 (42%)	323 (38%)	
Multidrug resistance ▲	239 (26%)	126 (28%)	157 (18%)	133 (16%)	<0.0001
Non-vaccine serotypes only (n)	143	60	415	631	
Penicillin	17 (12%)	5 (8%)	70 (17%)	140 (22%)	0.0066
pbp:1a-2b-2x†	17 (12%)	5 (8%)	70 (17%)	140 (22%)	
Erythromycin	1 (1%)	0 (0%)	12 (3%)	37 (6%)	0.032
ermB or *ermB/mefA*	1 (1%)	0 (0%)	7 (2%)	27 (4%)	
mefA	0 (0%)	0 (0%)	5 (1%)	10 (2%)	
Clindamycin	1 (1%)	0 (0%)	7 (2%)	27 (4%)	0.07
ermB or *ermB/mefA*	1 (1%)	0 (0%)	7 (2%)	27 (4%)	
Chloramphenicol	5 (3%)	2 (3%)	30 (7%)	31 (5%)	0.49
cat	5 (3%)	2 (3%)	30 (7%)	31 (5%)	
Tetracycline	8 (6%)	3 (5%)	41 (10%)	77 (12%)	0.026
*tet*M	8 (6%)	3 (5%)	41 (10%)	75 (12%)	
*tet*S/M	0 (0%)	0 (0%)	0 (0%)	2 (0.3%)	
Cotrimoxazole	65 (45%)	26 (43%)	212 (51%)	315 (50%)	0.35
*folA*_I100L	1 (1%)	0 (0%)	1 (0%)	4 (1%)	
*folP*_aa_insert_57-70	36 (25%)	16 (27%)	94 (23%)	123 (19%)	
*folA*_I100L, *folP*_aa_insert_57-70	28 (20%)	10 (17%)	117 (28%)	188 (30%)	

Multidrug resistance ▲	6 (4%)	2 (3%)	39 (9%)	73 (12%)	0.012

* Vaccine periods of interest for generalized linear regression analysis.

** A generalized linear regression model to detect a correlation of significance between the introduction of PCV and changes in antibiotic nonsusceptibility prevalence, comparing the pre-PCV and late-PCV13 periods.

† mutations in *pbp*1a, *pbp*2b, *pbp*2x, or a combination of mutations that confer penicillin nonsusceptibility.

▲ Multidrug resistance was defined as nonsusceptibility to three or more classes of antibiotics.

3.4. Discussion

Our study captures the genomic population structure and temporal trends among invasive pneumococci in South African children across a 16-year period which spans pre- and post-PCV eras. We highlight significant reductions in globally disseminated vaccine-type GPSC lineages following the introduction of PCV7 and PCV13. We show that the persistence of IPD after PCV13 introduction is, in part, attributed to both between and within lineage expansion of NVTs. Through diversity analysis, we show supporting evidence of pneumococcal population perturbation after vaccine introduction which reached equilibrium 11 years after the first use of PCV in 2009. We also show that although the prevalence of antibiotic-nonsusceptibility decreased overall, significant increases in nonsusceptibility to penicillin, erythromycin, and MDR were observed in IPD due to NVTs expressed by both vaccine- and non-vaccine-type lineages in the late-PCV13 period.

The decline in lineages predominantly expressing PCV13 serotypes is consistent with earlier epidemiological trend studies evaluating the impact of PCVs in South Africa, the USA and the UK [22,47,48,154–156]. However, similar to South Africa, vaccine-type lineages such GPSC5, GPSC9, GPSC10, and GPSC17 continue to cause invasive disease in the USA and the UK [58,87,98,111,157]. The GPS project database of 21 155 genomes from 59 countries revealed that three of the ten most common disease-causing lineages (GPSC17, GPSC21 and GPSC41) among South African children are endemic to the African continent, each with a contribution of at least 90% in the GPS collection (https://www.pneumogen.net/gps/gps-database-overview/). These lineages, particularly GPSC17 and GPSC21 are important drivers of residual VT disease in South Africa expressing VT 19A and 19F, respectively [158]. Serotype 19A is a genetically diverse serotype that was a major cause of IPD in children in

the USA after PCV7 introduction, and earlier studies identified antimicrobial-resistant emergent genotypes that expanded due to positive antibiotic selective pressure [97,159]. For example, the MDR CC320/271 clone outcompeted the antimicrobial susceptible clone CC199 [97]. Similar findings were reported in Canada [158], and some Asian countries [160]. By contrast, serotype 19A isolates from South Africa were historically represented by the CC2062 (GPSC17) clone [100], and this remains the predominant clone after PCV introduction. Similarly, the persistence of serotype 3 IPD was driven by ST458 (GPSC51), a clone that is predominant in South Africa but rarely reported elsewhere [100,101]. GPSC17 (CC2062), GPSC21 (CC347) and GPSC51 (CC458) require continued genomic surveillance in South Africa and other African countries as these lineages express vaccine-escape serotypes [156,157]. The persistence of serotype 3, however, was not surprising as PCV13 has been shown to have reduced vaccine effectiveness against this serotype [156]. Serotype 3 continues to have a steady incidence in South Africa particularly in the elderly where serotype replacement was evident [47]. As expected, vaccine introduction in South Africa was accompanied by significant increases in serotype diversity in the early-PCV13 period followed by declines towards baseline trends in the late-PCV13 period. Pooled serotype and lineage diversity trends followed similar patterns supporting the phenomenon of expected increases in the diversity of the pneumococcal population following the perturbation of genetic composition due to vaccine introduction [143,161].

Apart from residual VT IPD, NVTs play an important role in the persistence of invasive disease in children globally in the post-PCV era. Genomic surveillance of all serotypes is useful for understanding the genetic background of lineages driving persistence of NVT disease. We observed NVT IPD due to both within and between lineage expansions. Among the ten most common lineages in our collection, non-vaccine-type GPSC3 (predominantly

expressing serotype 8) substantially contributed to IPD in children after vaccine introduction. Before PCV7 introduction in South Africa, serotype 8 was the second most common NVT causing IPD in children aged <2 years (after serotype 15B/C) [100]. After PCV13 introduction, serotype 8 became the most common cause of IPD in both children and adults [162], moving from a frequency of 6% in the early-PCV13 period to 17% in the late-PCV13 period (data not shown). In England and Wales, serotype 8 IPD frequency increased substantially and accounted for 20% of disease after PCV13 introduction [141]. NVT 12F (predominantly expressed by GPSC26 and GPSC56), 15B/C (GPSC25 and GPSC48) and 16F (GPSC33 and GPSC46) were also frequently isolated from children with IPD in the current study after PCV13 introduction. Previous data from South Africa showed that serotype 15B/C was significantly associated with increased in-hospital deaths and serotype 16F was most commonly isolated among HIV-infected children [154,162]. Serotype 8, 12F and 15B/C are included in higher-valency vaccines PCV20 and PCV24. The vaccine IVT-25 covers in addition serotype 16F with a trade-off of PCV13 serotype 6A.

Previous studies have reported on evidence of shared genetic backgrounds among isolates of different pneumococcal serotypes [105,163]. We observed within lineage NVT expansions among GPSC5, GPSC9 and GPSC10. The expansion of NVT 35B within GPSC5 in the late-PCV13 period in this study coincided with earlier findings where serotype 35B increases were accompanied by a four-fold increase in penicillin-nonsusceptibility in children aged <5 years [98]. According to the global antimicrobial use data, the use of broad-spectrum penicillins in South Africa increased substantially from a defined daily dose (DDD) of 2023 per 1000 population in the year 2000, to 3794 DDD in 2015 (https://resistancemap.onehealthtrust.org/). It is therefore reasonable to suggest that antibiotic selective pressure has contributed, in part, to the persistence of penicillin-nonsusceptible

lineages. Similar patterns of penicillin-nonsusceptible lineages were observed in Nepal and France for GPSC9 and GPSC10, respectively [87,164]. The historic population dynamics of GPSC9 and GPSC10 together with GPSC5 (serotype 35B) and GPSC17 (serotype 19A) suggest that penicillin-nonsusceptibility likely promoted the expansion of these lineages in the past. Declines were observed in the effective population size of these lineage around the time when penicillin was widely used in children to treat streptococcal infections in the late 1950s (https://resistancemap.onehealthtrust.org/AbxTimeline.php), but this was followed by an exponential increase as penicillin resistance emerged. In South Africa, pneumococcal infections are mainly treated with penicillin or ampicillin. The persistence of lineages that are nonsusceptible to these antibiotics may pose challenges in IPD management in our healthcare facilities.

The expansion and extinction of GPSC lineages underlie the trends in serotype incidence and AMR through the AMR genes these lineages harbour in disease. Identifying lineages that express NVTs that are increasing in incidence may be an early indication of which NVTs may be of concern in disease, particularly those that carry resistance genes against multiple antibiotic classes and require continuous monitoring. Before vaccine introduction in South Africa (2000–2005), PCV7 serotypes plus serotype 6A accounted for 94% of erythromycin-nonsusceptible isolates, with VT 14 accounting for 40% of invasive disease [165]. These findings correspond with the high *ermB* prevalence among serotype 14 isolates expressed by the GPSC10 (CC230) lineage, and NVT 15A isolates expressed by GPSC9 (CC63) in the current study. Macrolide-resistant GPSC10 (CC230, serotype 14 or Denmark^{14}-32) is an MDR lineage that continues to cause disease in South Africa in the post-PCV13 era and is responsible for the global spread of MDR serotype 24F [87]. Although the expansion of NVT 15A (GPSC9) and serogroup 24 (GPSC10) after PCV introduction in South Africa does not

outweigh the reduction of VT 14 expressed by both lineages, the time-calibrated phylogenies of these lineages suggest that these NVTs diverged long before PCV introduction and successfully propagated to current times when expansion was facilitated by the removal of VT 14. The high prevalence of cotrimoxazole resistance was not surprising since this antibiotic was widely used as prophylaxis in HIV-infected populations before routine antiretroviral therapy in the early 2000s, in South Africa [166].

This study included approximately 20% of viable IPD isolates from individuals aged <18 years, with sampling bias towards younger children aged ≤2 years. Thus, our findings may not accurately capture the vaccine impact among older children, especially the 5–17 year olds. In addition, sampling limits our ability to confidently identify all emerging genotypes. Therefore, we may have missed detecting earlier genomic changes which may lead to an underestimation of circulating lineages. Our AMR findings may also not fully reflect resistance trends due to the usage of a sampled study population. The dichotomization of vaccine-type and non-vaccine-type lineages may be sensitive, partly due to geotemporal differences in serotype distributions. Due to the inclusion of only IPD isolates, interpretation of our phylodynamic analysis was limited to effective population size trends and did not account for within-host pathogen evolution through asymptomatic colonization. Nonetheless, the impact of vaccination may be more accurately reflected in the ≤2 years as this is the target age group for vaccination given the high disease burden. Our study collection was a snapshot of the GERMS-SA database across all four vaccine periods, hence the expectation that drawn conclusions closely reflect those of the surveillance program. Although not fully assessed, we hypothesise that there is no systematic bias for cases diagnosed by PCR alone. The use of the standardized GPSC lineage typing scheme to evaluate vaccine impact proved useful as it offers global comparability at the genome level. Due to the high concordance among

phenotypic and in silico predictions, the use of WGS to enrich our surveillance through the routine sequencing of IPD isolates or culture-negative specimens is a possibility in our setting going forward.

In conclusion, pneumococcal genomic surveillance provides a powerful means to evaluate the impact of interventions with respect to genomic characteristics. In South Africa, starting in 2024, PCV13 will be replaced by the more affordable SII PCV10 (Serum Institute of India, Pneumosil) which excludes serotypes 3, 4, and 18C in the formulation. Although serotype 4 and 18C were less important in infants and children before PCV introduction, they were, however important in adult populations that will no longer benefit from the indirect protective effect against these serotypes. Serotype 3 remained largely unaffected by PCV13. Genomic surveillance in all age groups will be particularly important in our setting to monitor the impact of switching to a lower valency vaccine as well as monitoring non-vaccine serotype fluctuations, such as serotype 8. Collaborative genomic epidemiology efforts within the African continent are useful for the in-depth evaluation of endemic lineages that are contributing to both residual vaccine serotype and non-vaccine serotype disease. Not only will this effort improve geographical representation, it may also contribute towards future vaccine designs and inform policy making within the African countries.

Chapter Four - Trends in antibiotic resistant invasive pneumococcal disease caused by non-13-valent pneumococcal conjugate vaccine serotypes and lineages in South Africa

4. Chapter 4 reference:

Lekhuleni C, Ndlangisa K, Ferreira AD, Cheng HR, Lorenz O, de Gouveia L, Quan V, Hawkins PA, McGee L, Bentley SD, Meiring S, Cohen C, Lo SW, du Plessis M, von Gottberg A. Trends in antibiotic resistant invasive pneumococcal disease caused by non-13-valent pneumococcal conjugate vaccine serotypes and lineages in South Africa, 2005–2020. *In preparation.*

4.1. Introduction

AMR has emerged as a global threat to health in the 21st century [167] and is predicted to cause an estimated 10 million deaths annually by 2050 in the absence of policies and interventions to curb AMR spread [168]. In 2019, an estimated 5 million deaths were associated with bacterial AMR [169]. *S. pneumoniae* is an important contributor to AMR and is included in the 2024 WHO Bacterial Priority Pathogens List [170,171]. Vaccines are instrumental in reducing antibiotic-resistant bacterial infections [172,173]. After the introduction of PCV, several countries, including South Africa, reported reductions in overall antibiotic-resistant IPD due to PCV targeting serotypes that were largely contributing to resistant IPD [47,89]. However, PCV use is often followed by increases in the prevalence of

NVTs [90]. This change in serotype distribution is often accompanied by increases in NVT antibiotic-resistant IPD [90].

Genotyping of resistant strains using MLST has shown that MDR pneumococci belonged to a small number of globally disseminated clones defined by the PMEN network [174,175]. PCV7 targeted the majority of pre-PCV antibiotic-resistant serotypes expressed by these clones, especially β-lactam-resistant isolates in the USA [176]. After PCV introduction, several countries reported serotype replacement [58]. For example, following PCV13 introduction in the USA, the Spain^{9V}-3 clone (PMEN3) drove the expansion of MDR NVT 35B through multiple successful capsular switching events from VTs 9V, 14, and 19A [59,70]. Several countries have shown that GPSCs that express both VT and NVT variants, which are sometimes capsular switch variants, tend to persist after PCV introduction through the expansion and diversification of NVT clones within these lineages [111,112]. Our study aimed to evaluate trends in antibiotic-nonsusceptible IPD and to identify MLST genotypes and GPSC lineages driving pneumococcal AMR in South Africa after PCV13 introduction

4.2. Methods and materials

4.2.1. Invasive pneumococcal disease surveillance in South Africa

Invasive disease surveillance was as described in section 2.1.

4.2.2. Phenotypic characterisation

Phenotypic characterisation was as described in section 2.2.

In addition, for this analysis VT was defined as serotype contained in the PCV13 formulation. NVT was defined as serotype not targeted by PCV13. Unless otherwise stated, serotype refers to phenotypic serotype. We defined penicillin-MDR as nonsusceptibility to penicillin plus two or more of the following antibiotics: erythromycin, chloramphenicol, tetracycline, rifampicin, and sulfamethoxazole-trimethoprim (cotrimoxazole). Other MDR was defined as nonsusceptibility to three or more of the above antibiotics, excluding penicillin. For all antibiotics, intermediately-resistant and fully-resistant isolates were grouped as nonsusceptible isolates.

4.2.3. Whole-genome sequencing and molecular characterisation

WGS and genomic data processing are as described in sections 2.3–2.5. In this chapter, STs are only mentioned if they do not belong to any CC. See look-up table in Appendix B.

4.2.4. Statistical analysis

IPD incidence was determined using all IPD cases from South Africa, 2005–2020. We defined PCV periods as: pre-PCV (2005–2008), PCV7 (2009–2010), early-PCV13 (2011–2014), and late-PCV13 (2015–2020). We determined IPD incidence (per 100,000 population) for age groups: <5 and ≥5 years by comparing the average number of cases in the pre-PCV period with those in the late-PCV13 period. The chi-squared test was used to evaluate the frequencies of nonsusceptible NVT IPD comparing the pre-PCV and late-PCV13 periods. The two-sample proportion test was used to evaluate differences in nonsusceptibility profiles comparing the pre-PCV and late-PCV13 periods. Statistical analyses were performed in Stata 18.0 (StataCorp, College Station, Texas, USA) and R version 4.2.1.

4.3. Results

4.3.1. Invasive pneumococcal disease surveillance, South Africa

From 2005 through 2020, 54,199 IPD cases were reported to the NICD reference laboratory. Age data were available for 51,917 (95.8%) of the cases. Among cases with known age, 35,416 (68.2%) had viable isolates for phenotypic characterisation. Case characteristics by age category are shown in Table 4.1. Case characteristics including those with age data not captured are shown in Appendix L. Of the 8722 viable isolates from children <5 years, penicillin plus cotrimoxazole (penicillin-cotrimoxazole) nonsusceptibility was the most common phenotype (25.7%, n=2245), followed by penicillin-MDR (22.0%, n=1919). Similar to the <5 year olds, penicillin-cotrimoxazole nonsusceptibility was the most predominant phenotype among the 26,694 viable isolates from individuals ≥5 years (16.1%, n=4310), followed by cotrimoxazole-only nonsusceptibility (13.8%, n=3694). The incidence of overall IPD declined across all ages following reductions in VT IPD after vaccine introduction, with the largest decline among children <5 years (Figure 4.1). NVT IPD incidence remained relatively low and stable across the study period (Figure 4.1). Among NVTs, the overall proportion of isolates that were nonsusceptible to penicillin-cotrimoxazole, penicillin-MDR, and other-MDR increased from 3.6% (115/3227), 0.9% (30/3227), and 2.7% (87/3227) in the pre-PCV period to 8.4% (469/5586), 6.4% (356/5586), and 5.0% (280/5586) in the late-PCV13 period ($p<0.001$), respectively (Appendix M).

4.3.2. Trends in antibiotic-nonsusceptible invasive disease

In children <5 years, the overall incidence of penicillin-cotrimoxazole and penicillin-MDR IPD declined from 5.6 to 0.8 per 100,000 population (percentage decrease –86.5%, CI –90.5% to –81.4%) and 5.5 to 0.4 (–93.0%, –95.7% to –89.1%), respectively, when comparing

the late-PCV13 period to the pre-PCV period (Figure 4.2). Similar reduction trends were observed for PCV13-serotype-IPD. The incidence of nonsusceptible IPD due to non-PCV13 serotypes fluctuated across the study period (Figure 4.2). However, increases in the incidence of IPD nonsusceptible to cotrimoxazole were observed among children from 2012–2014 (Figure 4.2), partly due to cotrimoxazole-nonsusceptible NVT 15B/C that increased from a proportion of 1.7% (11/658) in the pre-PCV period to 6.0% (44/734, p<0.001) in the early-PCV13 period. Similarly, penicillin-cotrimoxazole nonsusceptible IPD due to NVT 15A and 35B increased from 0.2% (1/658) and 0.2% (1/658) in the pre-PCV period to 1.6% (12/734, p=0.01) and 5.0% (37/734, p<0.001) in the early-PCV13, respectively.

Among individuals ≥5 years, overall penicillin-cotrimoxazole and penicillin-MDR IPD declined from 0.7 to 0.4 (–48.9%, –57.7% to –38.6%) and 0.6 to 0.2 (–60.0%, –68.0% to –50.2%), respectively, when comparing pre-PCV and late-PCV13 periods. NVT penicillin-cotrimoxazole nonsusceptible IPD increased from 0.04 to 0.1 (185.4%, 60.9 to 434.2%) when comparing the same periods. Similar to the <5 year olds, cotrimoxazole nonsusceptibility was an important phenotype in the ≥5 year olds (Figure 4.2).

4.3.3. Residual PCV13 and non-PCV13 serotype antimicrobial nonsusceptibility among viable isolates

Among children <5 years of age, residual PCV13 IPD contributed 43.1% (112/260) and 52.3% (68/130) of penicillin-cotrimoxazole nonsusceptible IPD and penicillin-MDR IPD in the late-PCV13 period, respectively (Appendix M). Among the ≥5 year olds, PCV13 serotypes accounted for 71.2% (794/1115) and 59.8% (437/731) of penicillin-cotrimoxazole nonsusceptible IPD and penicillin-MDR IPD in the same period, respectively (Appendix M).

Overall, PCV13 serotypes 19A and 19F were the most common cause of penicillin nonsusceptibility in the late-PCV13 period (Figure 4.3), both contributed to 89.5% (811/906) and 32.9% (166/505) of penicillin-cotrimoxazole nonsusceptible IPD and penicillin-MDR IPD in the late-PCV13 period, respectively. We observed increases in the number of non-PCV13 serotypes that caused nonsusceptible-IPD across the four vaccine periods (Figure 4.3). Among all non-PCV13 serotypes, NVT 15A, 15B/C, 35B, 10A, 16F, and 23B (in order of prevalence) contributed 61.9% (545/881) of penicillin-nonsusceptible IPD in the late-PCV13 period (Figure 4.3). NVTs 15A, 35B, and 23B showed increases in MICs in the late-PCV13 period compared to the pre-PCV period (Appendix N).

4.3.4. Genomic features of antibiotic-resistant non-PCV13 serotypes using sampled data

Among 4669 genomes, the concordance between phenotypic and genomic inference of resistance for penicillin, erythromycin, chloramphenicol, tetracycline, and cotrimoxazole was 96.1%, 98.8%, 99.2%, 97.8%, and 92.4%, respectively. Concordance included isolates that were on the cusp of susceptible and resistance and had one dilution difference. Due to the high concordance that was also comparable to the literature, no genomes were excluded. Of the 4669 genomes, 1824 (39.1%), 648 (13.9%), and 901 (19.3%) had determinants that conferred resistance to penicillin (*pbp1a-2b-2x* allele combinations that confer resistance), erythromycin (*ermB* and/or *mefA*), and MDR (penicillin and erythromycin resistance-determinants, *cat*, *tet*, and/or *folA*/*folP* substitutions), respectively. The ten most common lineages (each with n≥10 isolates) in order of prevalence that had penicillin-resistance determinants in the pre-PCV period with their corresponding predominant serotypes were GPSC17 (19A), GPSC10 (14), GPSC41 (6A), GPSC5 (23F), GPSC14 (23F), GPSC1 (19F), GPSC18 (14), GPSC37 (6B), GPSC77 (6B), and GPSC9 (14) (Appendix O). These common

lineages accounted for 76.0% (443/583) of penicillin-resistant isolates and were all vaccine-type lineages. In the late-PCV13 period, GPSC17 (19A), GPSC10 (3, 10A, 14, and 24), GPSC5 (35B), GPSC9 (15A), GPSC21 (19F), GPSC1 (19F), GPSC48 (15B/C), GPSC14 (23F), GPSC168 (15A) and GPSC102 (23B) were the most common cause of penicillin-resistant IPD, accounting for 83.0% (431/519) of the isolates (Appendix P). Three of the ten GPSCs (GPSC48, GPSC102, and GPSC168) were non-vaccine-type lineages in this period.

Residual VT 19A and 19F were expressed by nine and 14 GPSCs, respectively, some of which had high genotypic diversity (Figure 4.5; Appendix Q). GPSC17 (VT 19A, CC2062) commonly caused penicillin-cotrimoxazole nonsusceptible IPD before and after PCV13 introduction (Figures 4.3, 4.4 and 4.5). GPSC1 (VT 19F, CC271) and GPSC21 (VT 19F, CC347) commonly caused penicillin-MDR and penicillin-cotrimoxazole IPD before PCV introduction. After PCV introduction, these lineages continued to cause nonsusceptible-IPD, with GPSC21 predominating from 2018–2020 (Figures 4.3, 4.4 and 4.5). Resistance profiles are shown in Appendix Q.

Similar to VTs 19A and 19F, we observed variations in the number of GPSCs expressing each of the six non-PCV13 serotypes that commonly caused penicillin-nonsusceptible IPD in the late-PCV13 period (Figures 4.6 and 4.7; Appendix Q). NVTs 15A and 15B/C were expressed by 15 lineages each and NVT 16F by only five lineages. Among NVT 15A isolates, vaccine-type GPSC9 (CC63 and ST2613) and non-vaccine-type GPSC168 (ST11766) commonly caused penicillin-MDR and penicillin-cotrimoxazole nonsusceptible-IPD, respectively, during the early- and late-PCV13 periods (Figures 4.6 and 4.7; Appendix Q). Among NVT 15B/C isolates, non-vaccine-type GPSC48 (ST7052) commonly

caused penicillin-cotrimoxazole IPD (Figure 4.7 and Appendix Q). Other GPSCs that were important causes of NVT penicillin-nonsusceptible IPD included vaccine-type GPSC10 (NVT 10A and serogroup 24, CC230), vaccine-type GPSC5 (NVT 23B, ST1349; NVT35B, CC172) and non-vaccine-type GPSC102 (23B, CC4423). NVT 16F was predominated by fully-susceptible or cotrimoxazole-nonsusceptible phenotypes (Appendix Q). Non-vaccine-type GPSC33 (16F, CC4088), GPSC36 (10A, CC2068), and GPSC46 (16F, CC3450) were important causes of nonsusceptible NVT IPD after PCV introduction. Overall, lineages that expressed both VTs and NVTs were more frequent than VT-only and NVT-only lineages in the late-PCV13 period (Appendix R).

Table 4.1: Characteristics of invasive pneumococcal disease cases, South Africa, 2005–2020. N=51,917 (total number of reported cases with known age), n=35,416 (total number of viable isolates with known age), n_seq=4669 (total number of sequenced isolates).

	<5 years (n_viable=8722; n_seq=2680)			≥5 years (n_viable=26694; n_seq=1989)			Total (n_viable=35416; n_seq=4669)		
	n_viable	n_seq	% seq (n_seq/n_viable)	n_viable	n_seq	% seq (n_seq/n_viable)	n_viable	n_seq	% seq (n_seq/n_viable)
Year									
2005	1215	187	15.4	2297	98	4.3	3512	285	8.1
2006	1081	197	18.2	2197	97	4.4	3278	294	9.0
2007	1084	191	17.6	2106	131	6.2	3190	322	10.1
2008	1098	197	17.9	2079	96	4.6	3177	293	9.2
2009	1009	195	19.3	2271	92	4.1	3280	287	8.8
2010	650	198	30.5	2133	100	4.7	2783	298	10.7
2011	464	199	42.9	1880	95	5.1	2344	294	12.5
2012	356	206	57.9	1721	124	7.2	2077	330	15.9
2013	322	183	56.8	1551	79	5.1	1873	262	14.0
2014	301	188	62.5	1401	95	6.8	1702	283	16.6
2015	216	169	78.2	1437	127	8.8	1653	296	17.9
2016	231	169	73.2	1331	127	9.5	1562	296	19.0
2017	210	129	61.4	1307	154	11.8	1517	283	18.7
2018	175	34	19.4	1148	246	21.4	1323	280	21.2
2019	180	143	79.4	1193	138	11.6	1373	281	20.5
2020	130	95	73.1	642	190	29.6	772	285	36.9
Vaccine period									
Pre-PCV	4478	772	17.2	8679	422	4.9	13157	1194	9.1
PCV7	1659	393	23.7	4404	192	4.4	6063	585	9.6
Early-PCV13	1443	776	53.8	6553	393	6.0	7996	1169	14.6

Late-PCV13	1142	739	64.7	7058	982	13.9	8200	1721	21.0
Sex									
Female	3849	1169	30.4	13865	1000	7.2	17714	2169	12.2
Male	4690	1468	31.3	12549	972	7.7	17239	2440	14.2
Unknown	183	43	23.5	280	17	6.1	463	60	13.0
Specimen type									
Blood	5761	1807	31.4	15541	1210	7.8	21302	3017	14.2
Cerebrospinal fluid	2706	777	28.7	9164	660	7.2	11870	1437	12.1
Other	255	96	37.6	1989	119	6.0	2244	215	9.6
Serotype									
PCV7	3956	908	23.0	6532	417	6.4	10488	1325	12.6
PCV13 unique serotypes	2217	647	29.2	8680	572	6.6	10897	1219	11.2
Non-PCV serotypes	2549	1125	44.1	11482	1000	8.7	14031	2125	15.1
No isolate available	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Nonsusceptibility profile									
Fully susceptible	2523	861	34.1	13155	980	7.4	15678	1841	11.7
Penicillin NS only	189	23	12.2	267	21	7.9	456	44	9.6
Cotrimoxazole NS only	1320	442	33.5	3694	290	7.9	5014	732	14.6
Penicillin-cotrimoxazole NS	2245	749	33.4	4310	338	7.8	6555	1087	16.6
Penicillin MDR	1919	467	24.3	3125	212	6.8	5044	679	13.5
Other MDR	243	105	43.2	955	112	11.7	1198	217	18.1
Other resistance	283	33	11.7	1188	36	3.0	1471	69	4.7

n_viable - number of viable isolates with susceptibility data

n_seq - number of sequenced isolates

% seq - percentage of sequenced isolates (number of sequenced isolates divided by number of viable isolates with susceptibility data)

n_ns - number of nonsusceptible isolates

n_seq_ns - number of sequenced nonsusceptible isolates

% seq_ns - percentage of sequenced nonsusceptible isolates (number of sequenced nonsusceptible isolates divided by number of nonsusceptible isolates)

NS = nonsusceptible

Penicillin multidrug resistance (MDR) was defined as nonsusceptibility to penicillin and two or more other antibiotic classes (erythromycin, chloramphenicol, tetracycline, rifampicin and cotrimoxazole)

Other MDR was defined as nonsusceptibility to three or more antibiotic classes

For sequenced isolates, MDR does not include rifampicin as the resistance determinant is not included in the bioinformatics pipeline

Audit cases were grouped into a ‘no isolate available’ category together with cases identified on PCR only (culture-negative) or cases where the isolate was never sent from the diagnostic laboratory or lost viability during transit, laboratory processing or storage

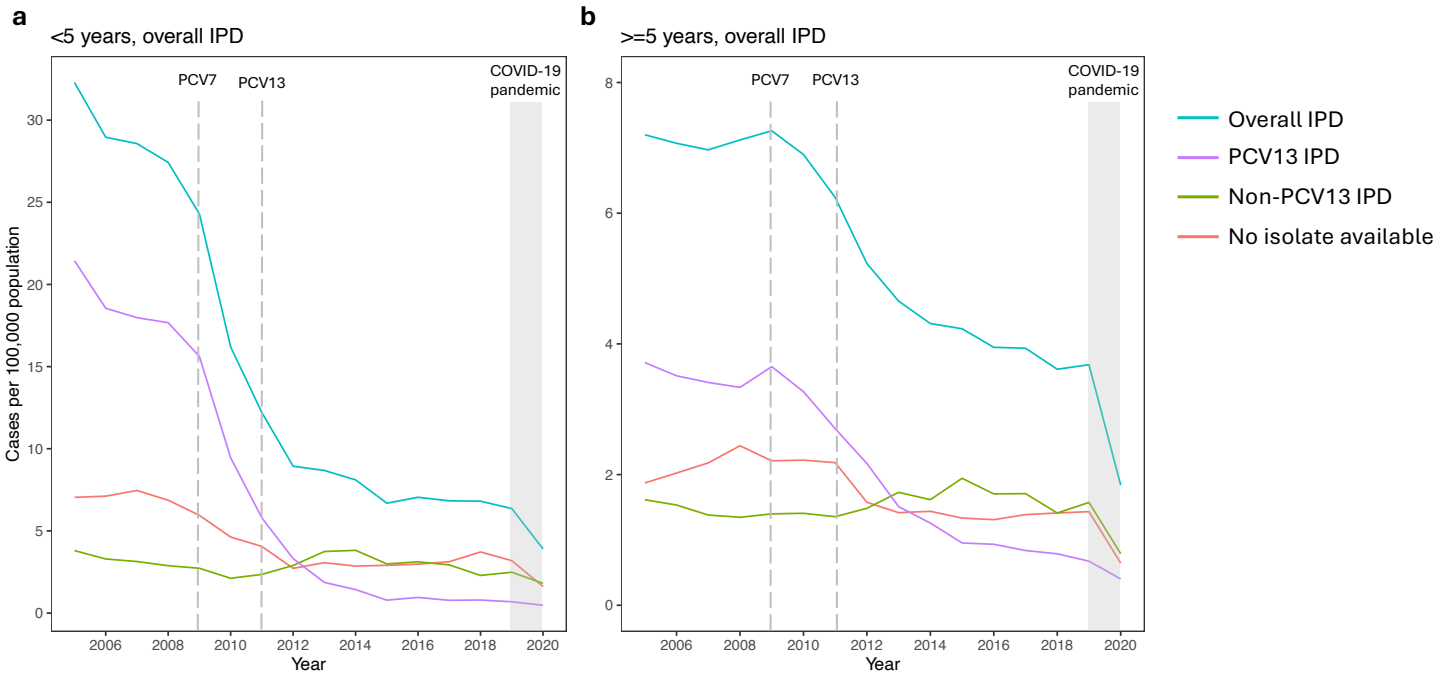


Figure 4.1: Incidence of reported invasive pneumococcal disease (IPD) in children <5 years (a) and individuals ≥5 years (b) in South Africa, 2005–2020. Incidence (per 100,000 populations) is shown by overall IPD, PCV13 IPD, non-PCV13 IPD, and cases that had no isolates available. PCV7 and PCV13 introduction years are shown by grey dotted lines. COVID-19 pandemic is highlighted by grey shading.

*No isolate available - include audit cases, culture negative specimens serotyped using PCR, missing and non-viable isolates.

*The y-axis scales are not standardised.

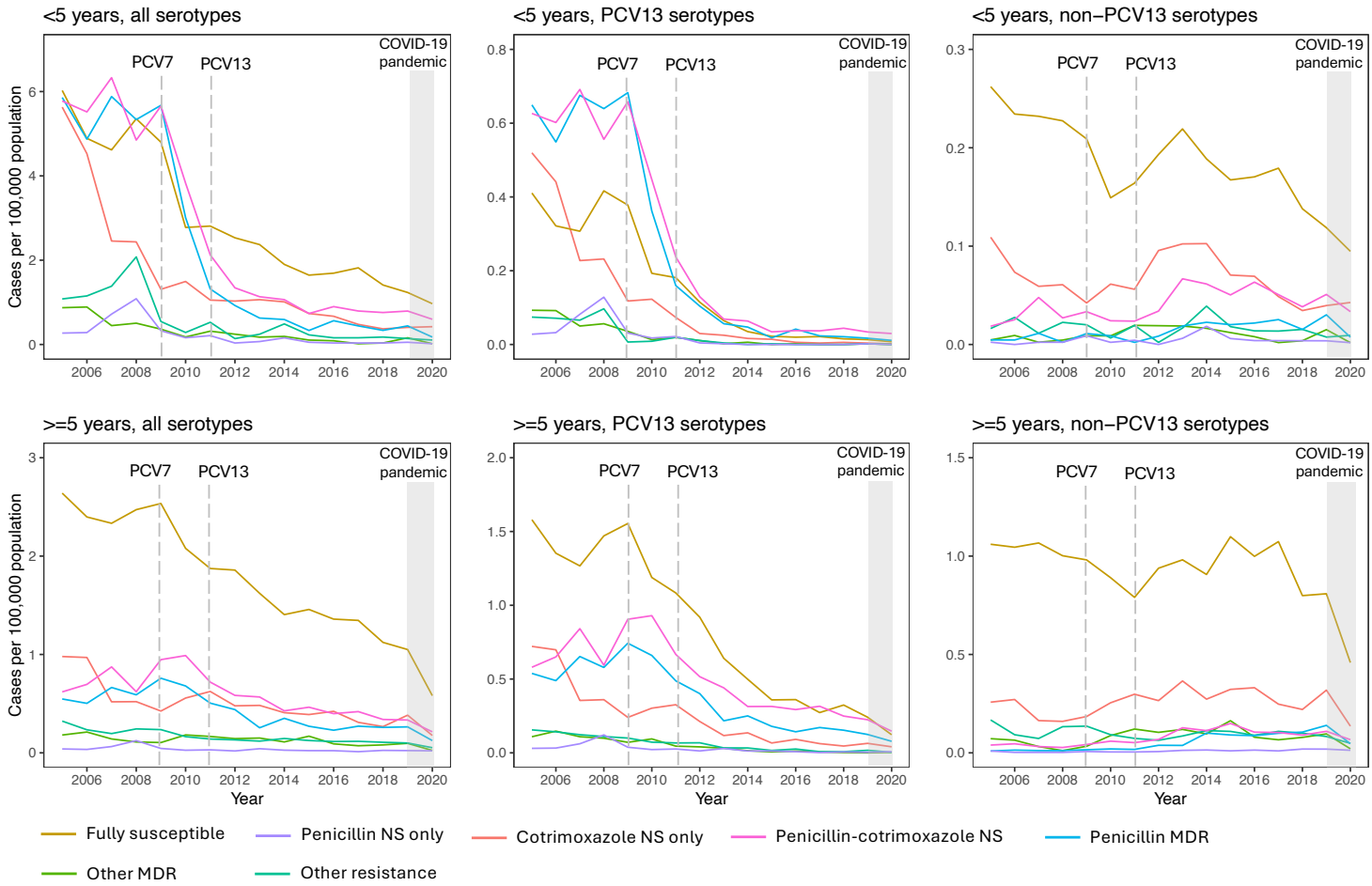


Figure 4.2: Incidence of nonsusceptible invasive pneumococcal disease (IPD) in children <5 years and individuals ≥5 years in South Africa, 2005–2020. Incidence (per 100,000 population) is shown by susceptibility profiles of the isolate (fully susceptible, penicillin-nonsusceptible (NS) only, cotrimoxazole NS only, penicillin-cotrimoxazole NS, penicillin-MDR, other MDR, and other resistance. PCV7 and PCV13 introduction years are shown by grey dotted lines. COVID-19 pandemic is shaded in grey.

*The y-axis scales are not standardised.

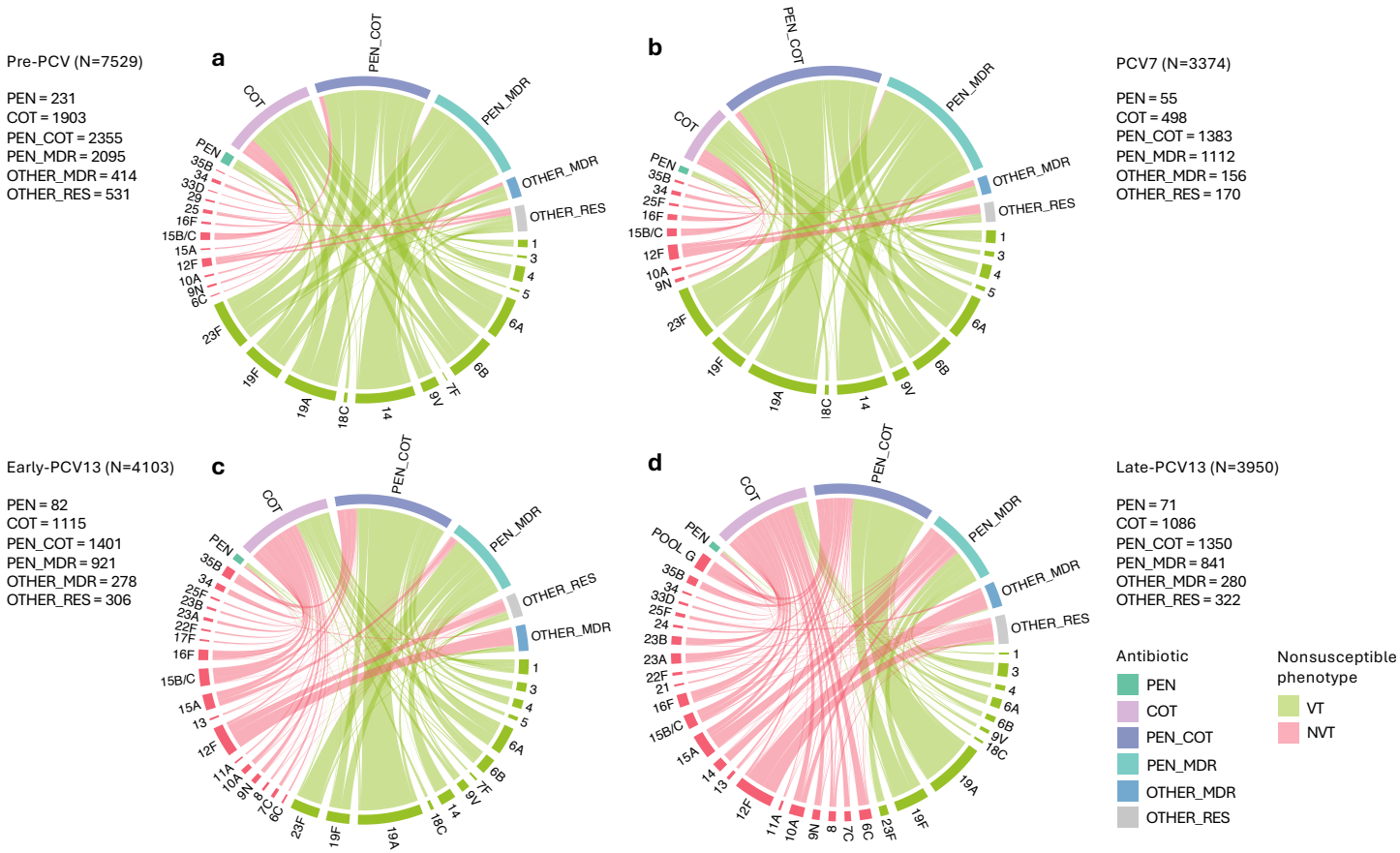


Figure 4.3: Antibiotic-nonsusceptible IPD by vaccine period in all ages, South Africa. a–d show nonsusceptibility trends in the pre-PCV (2005–2008), PCV7 (2009–2010), early-PCV13 (2011–2014), and late-PCV13 (2015–2020) periods. Nonsusceptibility categories are shown on the outer tracks of the figures. Nonsusceptibility due to PCV13 serotypes is shown by green inner track links, and due to non-PCV13 serotypes by red inner track links. PEN = penicillin-nonsusceptible (NS), COT = cotrimoxazole NS, PEN_COT = penicillin-cotrimoxazole NS, PEN_MDR = nonsusceptibility to penicillin plus two or more antibiotic classes, OTHER_MDR = nonsusceptibility to three or more antibiotic classes, OTHER_RES = nonsusceptibility to at most two of the following antibiotic classes: erythromycin, chloramphenicol, tetracycline, and rifampicin. We included serotypes that had ≥20 nonsusceptible isolates.

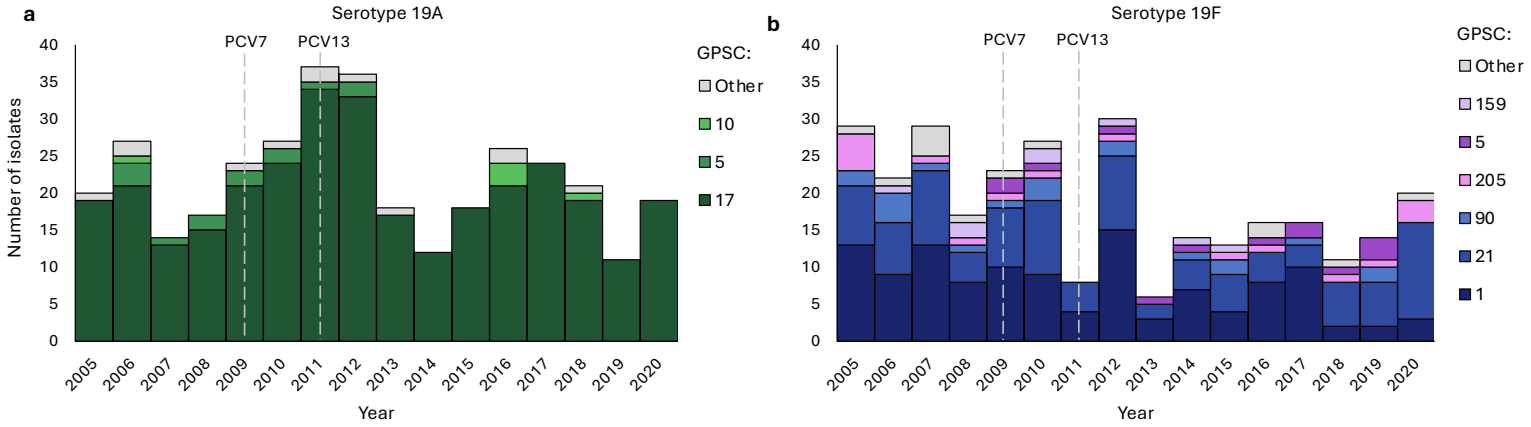
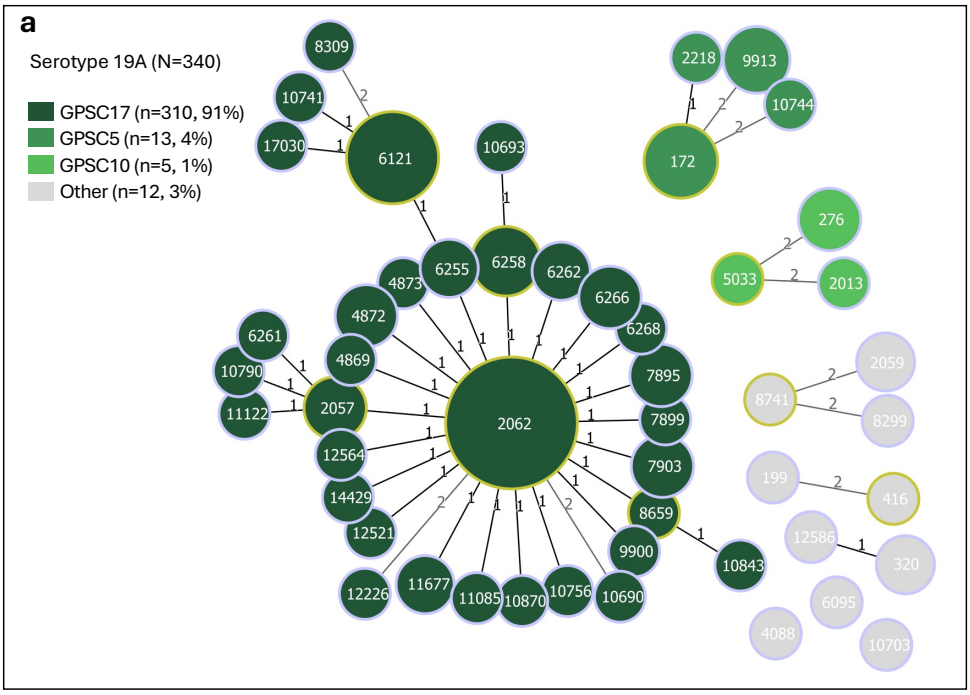


Figure 4.4: Global Pneumococcal Sequence Cluster (GPSC) lineages among residual PCV13 *in silico* serotypes 19A (a) and 19F (b) in South Africa, 2005–2020. PCV7 and PCV13 introduction years are shown by grey dotted lines.

* GPSCs that had <5 isolates were grouped into 'other'.

*Sampling after PCV13 introduction was biased towards non-PCV13 serotypes as the absolute number of sampled isolates remained unchanged even after VT disease declined.



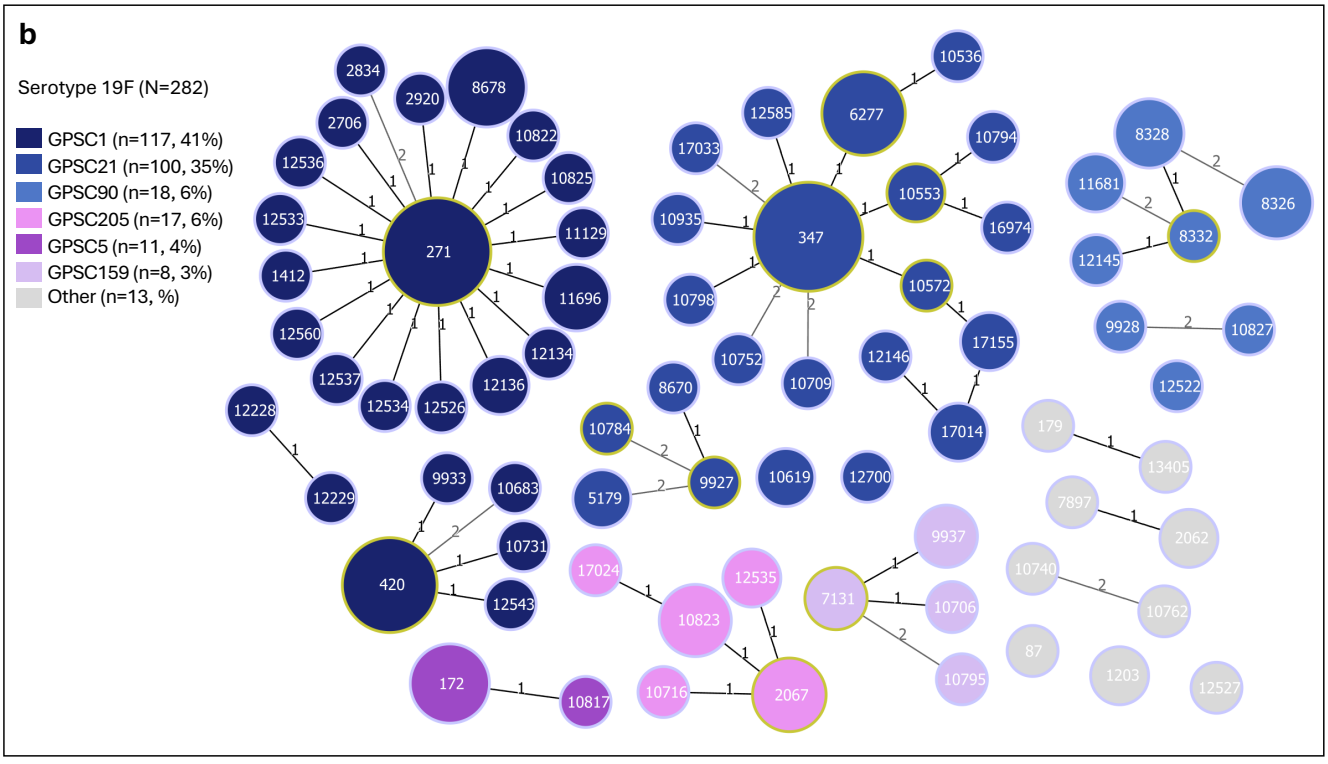


Figure 4.5: Multilocus sequence typing (MLST) minimum-spanning-trees of the two most common penicillin-nonsusceptible PCV13 serotypes 19A (a) and 19F (b) in South Africa in the late-PCV13 period. Genomic characteristics of isolates from 2005–2020 are included. Sequence types (ST) are coloured by Global Pneumococcal Sequence Cluster (GPSC) lineages. STs are linked by the number one if they shared

six of the seven MLST alleles, and by the number two if they shared five of the seven MLST alleles. STs that were different at three or more loci are not linked. Isolates with new and unknown STs are not shown.

* GPSCs that had <5 isolates were grouped into ‘other’ and coloured grey.

*Appendix Q shows the distribution of STs, GPSCs, and resistance phenotypes by vaccine periods.

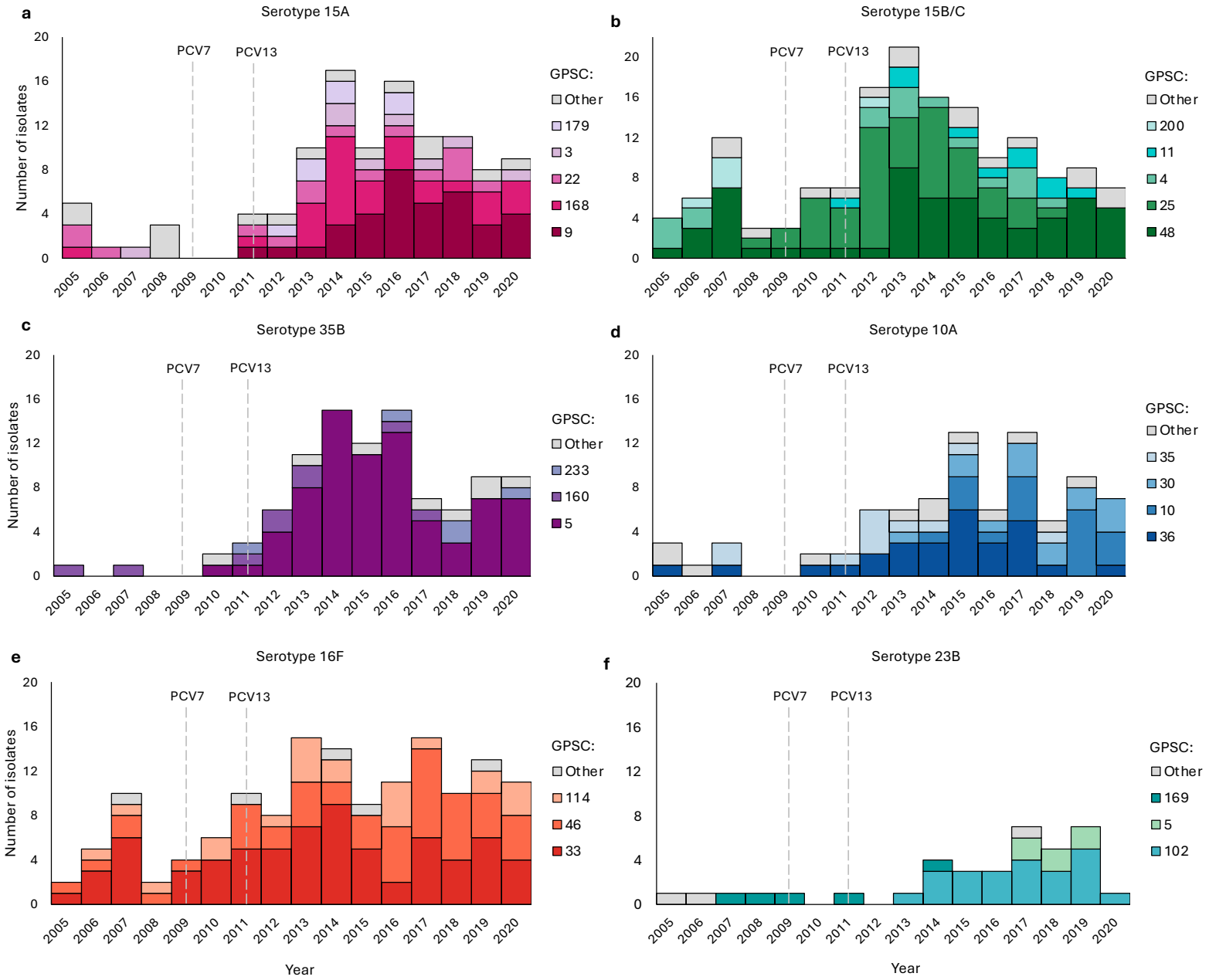


Figure 4.6: Global Pneumococcal Sequence Cluster (GPSC) lineages among the six (a-f) most common penicillin-nonsusceptible non-PCV13 serotypes in South Africa, 2005–2020. PCV7 and PCV13 introduction years are shown by grey dotted lines.

* GPSCs that had <5 isolates were grouped into 'other'.

*Sampling after PCV13 introduction was biased towards non-PCV13 serotypes as sampling strategy remained unchanged even after VT disease declined.

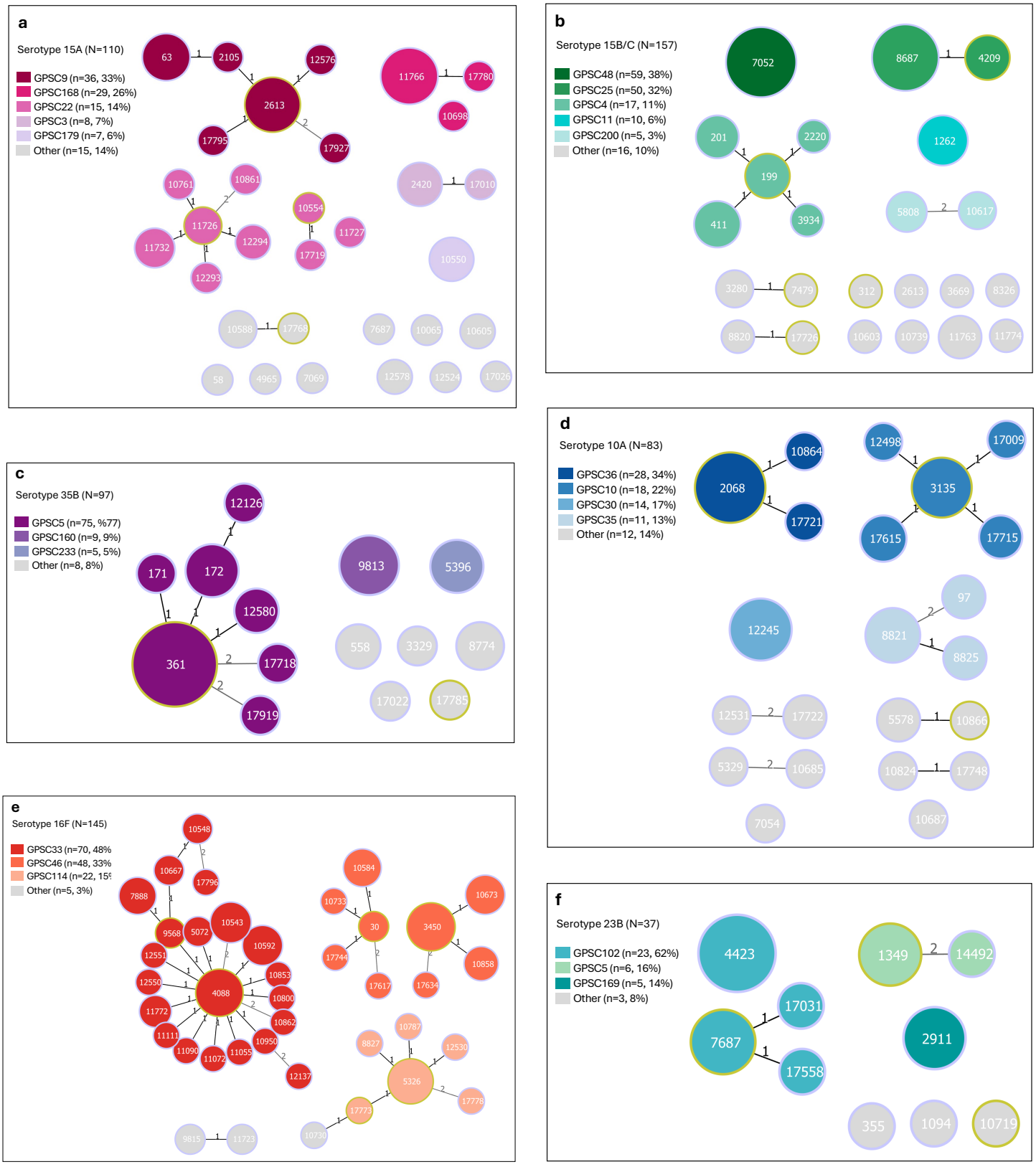


Figure 4.7: Multilocus sequence typing (MLST) minimum-spanning-trees of the six (a-f) most common penicillin-nonsusceptible non-PCV13 serotypes in South Africa in the late-

PCV13 period. Sequence types (ST) are coloured by Global Pneumococcal Sequence Cluster (GPSC) lineages. STs are linked by the number one if they shared six of the seven MLST alleles, and by the number two if they shared five of the seven MLST alleles. STs that were different at three or more loci are not linked. STs that have yellow borders are founder STs.

* GPSCs that had <5 isolates were grouped into ‘other’ and coloured grey.

*Appendix L shows the distribution of STs, GPSCs, and resistance phenotypes by vaccine periods.

4.4. Discussion

Our study, spanning a 16-year period, shows that following PCV introduction in South Africa, declines in the incidence of overall invasive disease due to reductions in PCV13-targeted serotypes led to a reduction in overall nonsusceptible IPD. Among antibiotic nonsusceptible IPD isolates, residual PCV13 serotypes 19A (predominantly CC2062) and 19F (CC347) contributed at least 40% of penicillin-nonsusceptible IPD in the late-PCV13 period. We show that although the incidence of NVT antibiotic-nonsusceptible IPD remained relatively stable, the number of NVTs that had antibiotic-nonsusceptible phenotypes increased across vaccine periods. Our findings also show that after PCV13 introduction, penicillin-nonsusceptible or penicillin-MDR disease due to residual VTs and common NVTs (15A, 15B/C, 35B, 10A, 16F, and 23B) was largely driven by lineages that were circulating before PCV introduction many of which showed high within-lineage genetic diversity.

Serotypes targeted by PCVs were major drivers of nonsusceptible-IPD before PCV use. The benefits of PCVs in reducing the burden of nonsusceptible-IPD were observed in South Africa (8,11). Our findings show that these reductions spanned multiple antibiotic classes in the nine years after PCV13 introduction. In the USA, the annual rate of nonsusceptible-IPD due to PCV13 serotypes declined across all age groups with a proportion reduction from 61% to 27% from 1998 through 2018 [85]. We report reductions of at least 90% of IPD nonsusceptible to penicillin plus other antibiotic classes in children younger than 5 years, and approximately 50% reductions in individuals ≥5 years. Similar VT nonsusceptible-IPD reductions for penicillin were reported in Hong Kong, [177], Israel [178], Malawi and The Gambia [58]; for macrolides in Iceland [179], Hong Kong, Israel, Malawi and The Gambia [58].

Although we observed significant declines in overall penicillin-nonsusceptible disease and MDR IPD after PCV introduction, residual PCV13 serotypes, especially serotype 19A and 19F continue to contribute substantially to penicillin-nonsusceptibility and MDR in the late-PCV13 period in our setting. After the widespread use of PCV7 (Pfizer, NY, USA) and PCV10 (Synflorix, GSK, Rixensart, Belgium), serotype 19A expanded and became the most common cause of IPD in some settings and became increasingly resistant to antibiotics, largely due to CC320 [180,181], a clone that emerged after PCV7 introduction and is related to the Taiwan^{19F}-14 (ST320) clone. While declines in serotype 19A IPD were reported after PCV13 introduction, more notably in younger children, several countries observed 19A increases, especially among adults [141]. Persistence of serotype 19F after PCV13 introduction was also reported in countries such as Portugal and England [182,183]. In South Africa, penicillin-nonsusceptible VT 19A persistence was due to CC2062, a clone that is uncommon in other settings but has been circulating for decades in South Africa and continues to cause disease [99,100]. In contrast, VT 19F persistence was driven by globally disseminated MDR CC271 (PMEN14) and penicillin-nonsusceptible CC347. Combined, these serotypes accounted for 42% of penicillin-nonsusceptible IPD between 2015 and 2020 in our study. Similar to other settings, vaccination of adults aged 19–64 years that have underlying medical conditions and the elderly (≥65 years), may further reduce resistant-VT invasive disease in South Africa [184].

In addition to residual nonsusceptible PCV13-serotype disease, the overall proportion of non-PCV13 serotypes that caused penicillin-nonsusceptible invasive disease increased from 5% before PCV introduction to 16% after PCV13 introduction among persons of all ages. In the 4

years after PCV13 introduction in South Africa, NVTs 10A, 15A, 15B/C, 16F, 23B, and 35B emerged and caused notable penicillin-nonsusceptible disease. These serotypes continued to expand from 2015 through 2020 and became the most common drivers of penicillin-nonsusceptible IPD among non-PCV13 serotypes. Not only did IPD frequency increase due to these NVTs, the minimum concentration required for effective penicillin treatment also increased notably among NVTs 15A, 23B, and 35B. In Botswana, increased pneumococcal colonisation by NVTs, especially 21 and 23B was observed following PCV13 use and was associated with invasive disease and AMR [185]. In England, a long-term upward trend in carriage and disease due to NVTs 10A, 15B/C, 23B, and 35B was observed 10 year after PCV13 introduction [183]. In the USA, NVTs 15A and 35B were among the most common causes of nonsusceptible-IPD after PCV13 introduction [85], with the latter showing elevated minimum inhibitory concentrations for penicillin and contributing substantially to β-lactam resistance driven by Utah^{35B}-24 that is characterised by MDR and global spread [70,85,87]. These observations were similar to those of the current study. Higher valency PCV20 that is recommended for use in some first-world countries may have the potential to reduce a proportion of IPD due to NVT 15B/C and 10A, the second and fourth highest contributors of penicillin-nonsusceptibility, respectively. In 2005 in South Africa, the defined daily dose (DDD) for broad spectrum penicillins was 1694 per 1000 population and increased to its highest in 2012, a year after PCV13 introduction to 5178 DDD per 1000 population (https://resistancemap.onehealthtrust.org/). This increasing penicillin use may have contributed to the ongoing antibiotic selective pressure favouring the emergence and expansion of these penicillin-nonsusceptible NVTs in the post-PCV13 era.

Previous studies have demonstrated the recombinogenic nature of some pneumococcal lineages [87,105]. Modelling has shown how vaccination could drive the expansion of pre-

existing resistant variants of NVTs following the removal of competing VT strains [90]. We observed changes in serotype composition within certain lineages after PCV introduction, similar to those observed in France where NVT 24F predominated the MDR GPSC10 (Denmark^{14}-32, (ST230)) lineage that was previously dominated by VT 14 pre-PCV [87]. In South Africa, MDR serogroup 24 was exclusively observed in the late-PCV13 period in this study. Similarly, seven NVTs predominated GPSC5 after PCV13 introduction in South Africa. In Germany, approximately 47% of NVT 23B and NVT 15A isolates were penicillin-nonsusceptible after PCV13 introduction, with ST1349-23B and ST63-15A being the most common cause of penicillin-nonsusceptible or penicillin-MDR IPD [186]. ST63-15A (GPSC9) has been reported with VTs 14 and 19A in Africa, Asia, Australia, and Europe, and has MDR resistance profiles. Therefore, historical capsular switching is a possibility within this lineage [111,186,187]. In South Africa, NVT 23B was common among children and NVT 15A was common in both children and adults after PCV13 introduction, similar to Germany [186].

Dissimilar to VT-NVT serotype composition changes post PCV, penicillin-nonsusceptible NVT 15B/C (GPSC48, ST7052) was already commonly circulating before vaccine introduction [100] compared to its variant VT 19F that was expressed by the same lineage and related clone (GPSC48, CC346) in this study. These findings are supported by the unchanged MIC_{50} of NVT 15B/C across vaccine periods, implying that serotype 15B/C fitness predated PCV introduction. We observed the expansion of NVT 16F (GPSC46) after PCV13 use which was largely driven by fully-susceptible and to a lesser extent, cotrimoxazole-nonsusceptible isolates. These findings corroborated those describing 16F epidemiology in Drakenstein, a community in the Western Cape province in South Africa where NVT 16F was the most commonly carried serotype with the expansion of

cotrimoxazole-nonsusceptible strains attributed to the use of cotrimoxazole as prophylactic treatment for HIV-exposed and infected individuals in the area [188,189]. Although predominantly circulating within the African continent according to the GPS database, NVT 16F (GPSC46) expansion was also observed in England in the post-PCV13 era and caused invasive disease [188,190].

Our study had limitations. Nonsusceptibility incidence rates in the late-PCV13 period were inclusive of 2020, a year that had low IPD cases due to non-pharmaceutical interventions introduced during the COVID-19 pandemic. We may therefore have underestimated the magnitude of incidence changes in this period compared to pre-PCV. However, statistically, there was no difference when excluding 2020 data. We used broad age groups (<5 and ≥5 years) which limited our ability to provide resolved nonsusceptibility incidence changes among age groups of importance in the literature such as those <2 years of age and individuals ≥65 years. However, overall nonsusceptibility trends among children and adults were similar to those of resolved age groups from South Africa and other settings. We sequenced only a subset of IPD isolates by year and may have underestimated the within-lineage serotype diversity before and after PCV introduction. In addition, NVTs were oversampled in the post-PCV period since our sampling strategy remained unchanged even after VT IPD declined. However, despite these limitations, our study captured the most important antibiotic nonsusceptible IPD serotypes and their genomic characteristics after PCV13 introduction in South Africa. These findings were comparable to other settings.

In conclusion, PCVs have been instrumental in reducing the burden of antibiotic-nonsusceptible IPD in South Africa. The emergence and expansion of nonsusceptible

serotypes that are not covered by PCVs may threaten this benefit if nonsusceptible NVT frequencies increase with time. Moreover, South Africa switched to the lower valency, more affordable PCV10 (Serum Institute of India, Pneumosil) in 2024 which warrants continued surveillance to monitor the impact of this change. Higher valency vaccines such as PCV20 that is currently licensed in the US and some European countries for use in high risk adults and the elderly may reduce antibiotic-nonsusceptibility due NVT 10A and 15B/C which are among the six most common contributors of penicillin nonsusceptibility after PCV13 introduction. NVT penicillin-nonsusceptible and MDR disease was driven by both vaccine-type lineages that predominantly expressed NVTs after PCV introduction and pre-existing non-vaccine-type lineages that expanded post PCV use.

Chapter Five - Clonal Expansion of Global Pneumococcal Sequence Cluster 3 within Serotype 8 after 13-Valent Pneumococcal Conjugate Vaccine Introduction, South Africa

5. Chapter 5 reference:

Lekhuleni C, Ndlangisa K, Ferreira ADS, Cheng HCR, Lorenz O, Kleynhans J, de Gouveia L, Skosana H, Quan V, Cremers AJH, Hawkins PA, Chochua S, McGee L, Bentley SD, Meiring S, Cohen C, Lo SW, von Gottberg A, du Plessis M. Clonal Expansion of Global Pneumococcal Sequence Cluster 3 within Serotype 8 after 13-Valent Pneumococcal Conjugate Vaccine Introduction, South Africa, 2005–2020. *In preparation.*

5.1. Introduction

Following PCV use, non-PCV13 IPD, especially adult-IPD due to serotype 8 increased substantially in countries such as Australia, Finland, France, and Norway [191]. In South Africa, non-PCV13 serotypes became the most common cause of IPD with serotype 8 being predominant, accounting for 12% of invasive disease across all ages between 2012 and 2018 [162]. By 2019, serotype 8 disease had increased notably from a proportion of 8% to 14% among individuals >64 years compared to the baseline period (2005–2008), with this increase partially offsetting the overall benefits of IPD reduction due to PCV use [47].

Before PCV introduction, ST53 was the predominant serotype 8 genotype in South Africa [100]. In Madrid, Spain, the ST53 clone of serotype 8 drove the increase in NVT invasive

disease after PCV13 introduction [192]. Similarly, in Denmark, serotype 8 ST53 was the dominant clone both before and after the introduction of PCVs and contributed to serotype 8 incidence increase in the elderly after PCV13 introduction [193]. Population snapshot studies conducted as part of the GPS project have shown that serotype 8 CC53 falls within the globally disseminated GPSC3 lineage [58]. In South Africa, GPSC3 was the most common cause of NVT invasive disease in children <3 years in the three years following PCV13 introduction (2011–2014) [58], and increased significantly (0.73 to 1.32 cases per 100 000 population, $p<0.008$) in children <5 years from the pre-PCV (2005–2008) to the late-PCV13 (2015–2020) periods [99]. Although the global contextualization of predominant lineages such as GPSC3 is essential for understanding lineage-specific invasive disease contributions, country-specific analyses are equally important for understanding localised expansion mechanisms and disease burden. Our study aimed to evaluate the expansion of serotype 8 in South Africa, and to describe and place circulating MLST genotypes and GPSC lineages into a global context.

5.2. Methods and materials

5.2.1. Invasive pneumococcal disease surveillance in South Africa

Invasive disease surveillance was as described in section 2.1.

5.2.2. Phenotypic characterisation

Phenotypic characterisation was as described in section 2.2.

5.2.3. Whole-genome sequencing and molecular characterisation

Serotype 8 isolates from South Africa were selected from the sampled collection described in section 2.3. WGS, genomic data processing and molecular characterisation are as described in section 2.4 and 2.5. Serotype 8 capsular locus sequences (CPS) were extracted using BLAST (version 2.15.0) and examined for genetic variations by *cps* alignment. The serotype 8 CPS reference CR931644 was used to evaluate gene content differences and to check for disruptive mutations to identify recombination events. Clinker was used for *cps* loci visualisation [194].

5.2.4. Phylogenetic analysis

A total of 686 serotype 8 genomes were analysed: 391 from South Africa (2005–2020) and 295 from 27 other countries in the GPS collection (1991–2020) by mapping against the *S. pneumoniae* reference genome ATCC 700669 (NCBI accession code FM211187) using Smalt version 0.7.4 (https://github.com/rcallahan/smalt). The pseudo-genome alignment was reduced to an alignment of variant sites using SNP-sites tool (version 2.5.1--hed695b0_0) [147]. The SNP alignment was used to construct a phylogenetic tree using the GTR model in FastTree2 (version 2.1.10--h470a237_2) [148]. A lineage-specific phylogeny of 749 GPSC3 isolates (including non-serotype 8 isolates of the same lineage) from South Africa (n=447, 2005–2020) and 14 other countries in the GPS collection (n=302, 1997–2020) was constructed by mapping Illumina reads against a GPSC3 reference (*Streptococcus pneumoniae* AP200 accession code CP002121) using BWA (version 0.7.17-r1188) [150]. Gubbins (version 3.2.1) was used to identify recombination events among serotypes expressed by GPSC3, and to construct a recombination-free phylogeny using the GTR model in RAxML [117,151]. Microreact and Phandango were used for visualisation [149,152].

5.2.5. Clonal expansion analysis

A recombination-free phylogeny of a subset (n=219) of serotype 8 CC53 genomes from South Africa that resulted in a significant temporal signal was used for Bayesian inference of ancestral dates using BactDating to perform root-to-tip regression [123]. The CaveDive R package that uses a coalescent-based model to reconstruct effective population size over time was then used to analyse the expansion dynamics of the CC53 clone using reversible-jump MCMC inference ran for 100 million MCMC iterations [125]. A burn-in of 20% was used. Carrying capacity was defined as the maximum population size of an expanding subpopulation [125].

5.2.6. Statistical analysis

Statistical analysis was performed using all reported serotype 8 cases from South Africa, 2005–2020. We defined PCV periods as: pre-PCV (2005–2008), PCV7 (2009–2010), early-PCV13 (2011–2014), and late-PCV13 (2015–2020). We determined serotype 8 IPD incidence (per 100,000 population) for age groups: 0–4, 5–14, 15–24, 25–44, 45–64, and >64 years by comparing the average number of cases in the pre-PCV period with those in the late-PCV13 period. GPSC3 incidence was calculated using imputed counts by assuming that the proportion of GPSC3 from the serotype 8 genomes from South Africa by year and by age group was similar to that of the total number of serotype 8 isolates from IPD surveillance. We used Wilcoxon rank sum test to compare GPSC3 mutation rate with pre-existing non-GPSC3 mutation rates previously determined by Gladstone and colleagues [105]. Statistical analyses were performed in Stata 18.0 (StataCorp, College Station, Texas, USA) and R version 4.2.1.

5.3. Results

5.3.1. National invasive pneumococcal disease surveillance, South Africa

A total of 54,199 IPD cases were reported from 2005 through 2020, of which 39,377 (72.7%) had serotyping data. Serotype 8 accounted for 3.3% (453/13,725), 3.0% (202/6657), 5.6% (521/9308), and 13.4% (1297/9687) of IPD cases during the pre-PCV, PCV7, early-PCV13, and late-PCV13 periods, respectively. Of the 2473 serotype 8 cases identified during the study period, age was recorded for 2429 (98.2%), of which 2170 (89.3%) had viable isolates. Of these, 391 (18.0%) isolates from persons of all ages were sequenced (Table 5.1).

Serotype 8 IPD incidence increased significantly in the late-PCV13 period among children <5 years (IRR 2.3, 95% CI 1.3–4.1), adults 45–64 years (2.1, 1.2–3.7), and adults >64 years (2.8, 1.1–8.3), compared to the pre-PCV period (Figure 5.1a). The majority of serotype 8 cases (310/523, 59.3%) among children <5 years were detected from CSF, corresponding to a higher proportion (267/344, 77.6%) of children with meningitis clinical syndrome. In contrast, the majority of serotype 8 cases among adults ≥25 years were detected from blood, which accounted for 84.4% (162/192) of cases in adults >64 years, correlating to a higher proportion (58/75, 77.3%) of lower respiratory tract infection and bacteraemia syndromes (Figure 5.1b).

5.3.2. Antimicrobial resistance

The majority of serotype 8 isolates were susceptible to penicillin (98.1%, MIC_{90}=0.03 μg ml^{-1}), erythromycin (99.2%, MIC_{90}=0.06 μg ml^{-1}), tetracycline (98.0%, MIC_{90}=2 μg ml^{-1}), and cotrimoxazole (95.7%, MIC_{90}=0.5 μg ml^{-1}). When comparing phenotypic AST results among the 391 sequenced isolates from South Africa with *in silico* AMR predictions, concordances

of 98.2% (n=384), 99.5% (n=389), 99.0% (n=387), and 96.2% (n=376) were observed for penicillin, erythromycin, tetracycline, and cotrimoxazole, respectively.

5.3.3. Serotype 8 genotypes and GPSC lineages, South Africa

Using seven-locus MLST, 384/391 (98.2%) serotype 8 isolates were grouped into five CCs. The remaining seven isolates were different at five or more MLST loci. CC53 was the dominant clonal complex (371/391, 94.9%). We identified one predominant GPSC, namely GPSC3, composed of CC53 (n=371, 94.9%) and CC1012 (n=2, 0.5%). The remaining 18 isolates were grouped into four GPSCs and accounting for 4.6% of the isolates (Figure 5.1c). We observed significant increases in the incidence of GPSC3 IPD among children <5 years (IRR 2.4, 95% CI 1.4–4.3) and individuals of all ages (IRR 1.7, 95% CI 1.3–2.1).

5.3.4. Global distribution of serotype 8 genotypes and GPSC lineages

Among 686 serotype 8 isolates (391 from South Africa and 295 from 27 countries; Appendix S), 640 isolates (93.3%) were grouped into seven CCs, with varying predominance by continent. The remaining 46 isolates had distinct STs that did not belong into any CC. CC53 was predominant in Africa (374/413, [90.6%]), Europe (188/203 [92.6%]) and South America (9/10 [90.0%]). CC1480 was mainly found in North America (14/19 [73.7%]) in this collection. In contrast, serotype 8 isolates from Asia were mainly ST4216 (n=8), ST6022 (n=8), CC10588 (n=4), and CC12793 (n=5), all contributing to 83.3% (25/30) of the isolates. Similarly, ST6748 and CC53 equally contributed to 90.9% (10/11) of genotypes from Oceania. CC53 accounted for 84.4% (579/686) of the global collection. Despite the global dissemination of serotype 8 (CC53), we observed a geographical structure within the CC53 clade (Figures 5.2a and 5.2b). CC53 isolates from South Africa clustered together in a clade,

within which three subclades (clades I through III) were identified (Figure 5.2b and 5.2c). Two ST53 isolates from Mozambique clustered within clade III (Figure 5.2c). Overall, the serotype 8 *cps* locus was largely conserved among the 686 serotype 8 isolates from the global collection, although isolates with >10 SNPs per gene were observed for three (*wzg*, *wchA* and *wzx*) of the 12 *cps* genes and were distributed across different specimen types, syndromes, CCs and GPSCs (Appendix T).

Globally, we identified nine GPSC lineages that expressed serotype 8 in this collection. GPSC3 (CC53 and CC1012) was the most predominant accounting for 84.7% (581/686) of the isolates and circulated on all continents, except Asia (Figure 5.2b, Figure 5.3 and Appendix U). GPSC98 (CC1480, CC3406, ST2894 and ST4202) was the second most common lineage (6.7%, n=47) and similar to GPSC3, we observed global dissemination, except in Asia. Serotype 8 isolates from Asia were predominantly expressed by GPSC224 (17/30, 56.7%), of which ST4216 and ST6022 equally comprised 16/17 (94.1%) of the isolates (Figure 5.2b and Appendix U). Each of GPSC22, GPSC32, GPSC277, GPSC336, GPSC694, and GPSC732 expressed at most ten isolates in this collection and accounted for 3.5% (24/686) of the isolates (Appendix U). GPSC3 (serotype 8) remained largely susceptible to penicillin, erythromycin, tetracycline and cotrimoxazole (Figure 5.3b and 5.3c).

5.3.5. Lineage-specific analysis - GPSC3

A global collection of 749 GPSC3 isolates expressed ten serotypes, including non-PCV13 serotypes 8, 11A, 15A, 22F, 31, 33A, and 33F (n=730), and PCV13 serotypes 3, 18C, and 23F (n=19) (Figure 5.3). Serotype 8 accounted for 77.6% (n=581) of the GPSC3 isolates.

While GPSC3 largely expressed serotype 8 in Africa and Europe at frequencies of 83.6% (376/450) and 93.1% (188/202), respectively, GPSC3 predominantly expressed serotype 33F (57.1%, 32/56) and 11A (30.4%, 17/56) in North America (Figure 5.3). There was no serotype 8 observation among the 16 GPSC3 isolates from Asia, however, serotype 11A (n=5) and 18C (n=9) accounted for 87.5% of the isolates, combined. GPSC3 isolates from South Africa (N=447) expressed all, except two of the aforementioned serotypes (18C and 23F). Comparative analysis of the *cps* gene clusters of serotypes expressed by GPSC3 showed that the serotype 8 *cps* locus was closely related to that of 15A and 11A (Figure 5.4). Individually annotated *cps* loci are shown Appendix V. Among the ten GPSC3 serotypes, serotype 8 was among the serotypes with minimal recombination events (Figure 5.5).

Similar to the serotype 8 global collection, *in silico* AMR predictions for GPSC3 (N=749) showed 100% susceptibility to penicillin (Figure 5.3). However, acquired resistance to erythromycin (3.9% [29/749]) conferred by *mefA* (86.2% [25/29]) and *ermB* (13.8% [4/29]), tetracycline resistance (2.1% [16/749]) conferred by *tet*M, and cotrimoxazole resistance (7.9% [59/749]) by the presence of mutation I100L in *folA* and/or indel within amino acid residue 56–67 in *folP* were observed, more notably among the non-serotype 8 isolates expressed by GPSC3 (Figure 5.3). Of the 16 isolates that had the *tet*M gene, 13 (81.3%) were borne on the repUS43 plasmid replicon.

5.3.6. Clonal expansion of CC53 (GPSC3) genomes from South Africa

The CC53 root-to-tip regression analysis showed a significant temporal signal (R^2=0.20 and p<0.001), with the most recent common ancestor occurring around 1929 (CI 1919–1937) (Appendix W). The 95% credible intervals for inferred ancestral dates are also shown in

Appendix W. CaveDive analysis estimated at least five clonal expansions within the CC53 isolates from South Africa (Figure 5.6a), all of which occurred before the introduction of PCV with the most recent expansion occurring around 1992 (CI 1990–1994). Predicted future population dynamics of clonal expansion branches showed steady increases in the effective population sizes of CC53 subpopulations over the next ten years, at least, in the absence of a serotype 8-targeting intervention (Figure 5.6b).

Table 5.1: Serotype 8 invasive pneumococcal disease isolates stratified by vaccine period (pre-PCV, PCV7, early-PCV13, and late-PCV13) among individuals of all ages in South Africa, 2005–2020, N=2429 (number of cases with known age), n=391 (number of sequenced isolates).

	Pre-PCV (2005-2008)		PCV7 (2009-2010)		Early-PCV13 (2011-2014)		Late-PCV13. (2015-2020)		Total	
	N	n seq. (%)	N	n seq. (%)	N	n seq. (%)	N	n seq. (%)	N	n seq. (%)
	439	30 (7)	197	9 (5)	504	83 (17)	1289	269 (21)	2429	391 (16)
Year										
2005	100	8 (8)								
2006	119	8 (7)								
2007	114	5 (4)								
2008	106	9 (9)								
2009			97	5 (5)						
2010			100	4 (4)						
2011					104	10 (10)				
2012					124	18 (15)				
2013					137	28 (20)				
2014					139	27 (19)				
2015							203	39 (19)		
2016							226	47 (21)		
2017							257	40 (16)		
2018							228	44 (19)		
2019							227	44 (19)		
2020							148	55 (37)		
Province										
Eastern Cape	20	2 (10)	10	0 (0)	38	9 (24)	120	24 (20)	188	35 (19)
Free State	23	0 (0)	16	0 (0)	30	7 (23)	62	17 (27)	131	24 (18)
Gauteng	243	17 (7)	100	7 (7)	244	43 (18)	472	107 (23)	1059	174 (16)
KwaZulu-Natal	35	0 (0)	15	0 (0)	47	10 (21)	67	13 (19)	164	23 (14)
Limpopo	6	0 (0)	6	0 (0)	6	1 (17)	58	13 (22)	76	14 (18)
Mpumalanga	12	1 (8)	5	0 (0)	18	4 (22)	51	9 (19)	86	14 (16)
Northern Cape	5	0 (0)	6	1 (17)	11	0 (0)	35	6 (17)	57	7 (12)
North West	21	1 (5)	3	0 (0)	27	2 (7)	41	10 (24)	92	13 (14)
Western Cape	56	9 (16)	35	1 (3)	82	7 (9)	374	70 (19)	547	87 (16)
Unknown	18	0 (0)	1	0 (0)	1	0 (0)	9	0 (0)	29	0 (0)
Sex										
Male	236	15 (6)	104	4 (4)	243	43 (18)	661	146 (22)	1244	208 (17)
Female	196	15 (8)	90	5 (6)	255	37 (15)	619	120 (19)	1660	177 (11)
Unknown	7	0 (0)	3	0 (0)	6	3 (50)	9	3 (33)	25	6 (24)
Age										
0-4 years	75	14 (19)	41	4 (10)	114	54 (47)	293	146 (50)	523	218 (42)
5-14 years	22	1 (5)	12	1 (8)	20	4 (20)	39	4 (10)	93	10 (11)
15-24 years	22	0 (0)	19	0 (0)	23	1 (4)	58	10 (17)	122	11 (9)
25-44 years	216	14 (7)	74	2 (3)	208	19 (9)	454	56 (12)	952	91 (10)

45-64 years	81	1 (1)	44	2 (5)	107	3 (3)	315	37 (12)	547	43 (8)
>64 years	23	0 (0)	7	0 (0)	32	2 (6)	130	16 (12)	192	18 (9)
Specimen type										
CSF	170	11 (7)	72	2 (3)	231	46 (20)	448	110 (26)	921	169 (18)
Blood	253	19 (8)	109	7 (6)	242	36 (15)	786	154 (20)	1390	216 (16)
Other	16	0 (0)	16	0 (0)	31	1 (3)	55	5 (9)	118	6 (5)

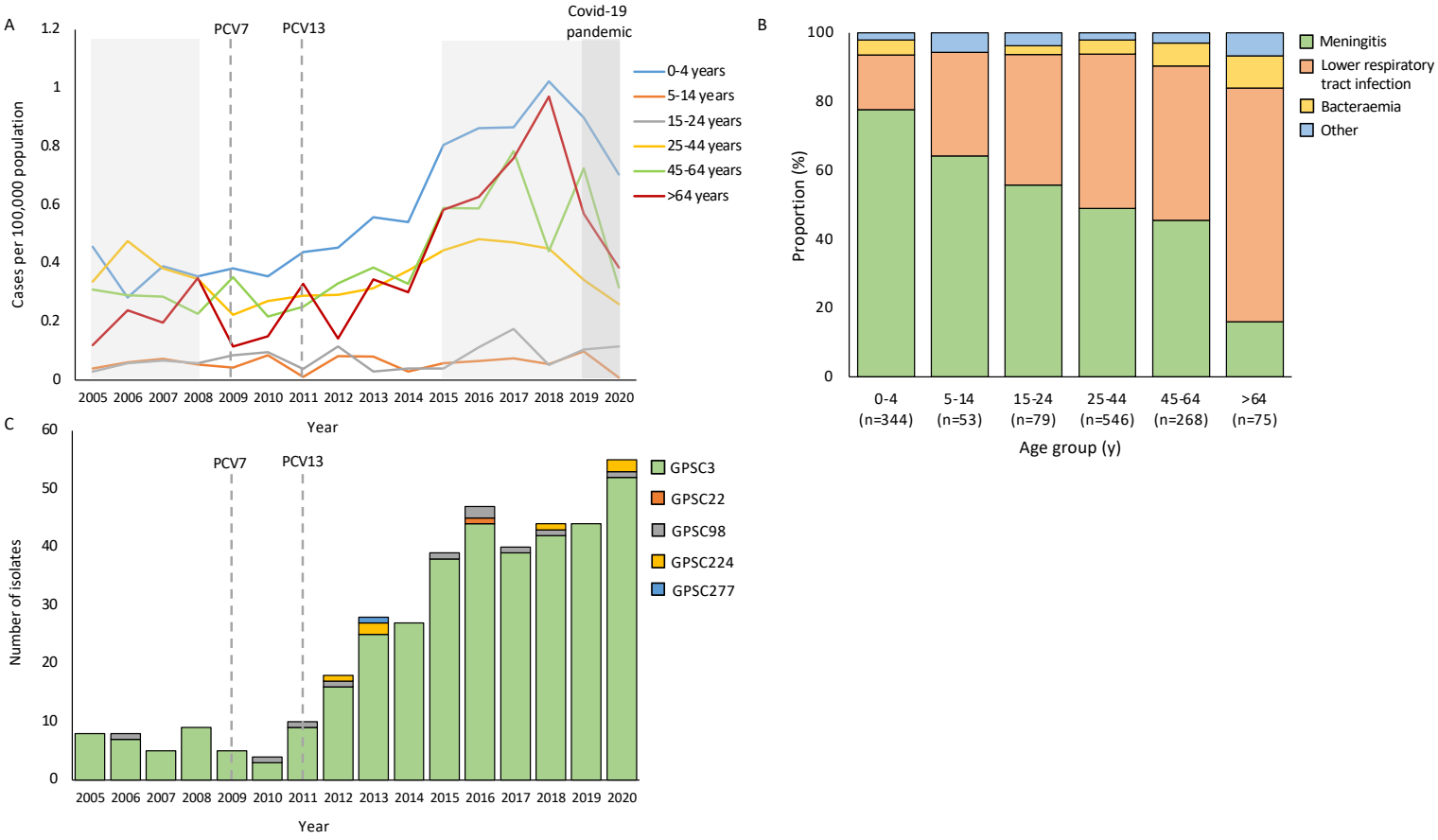
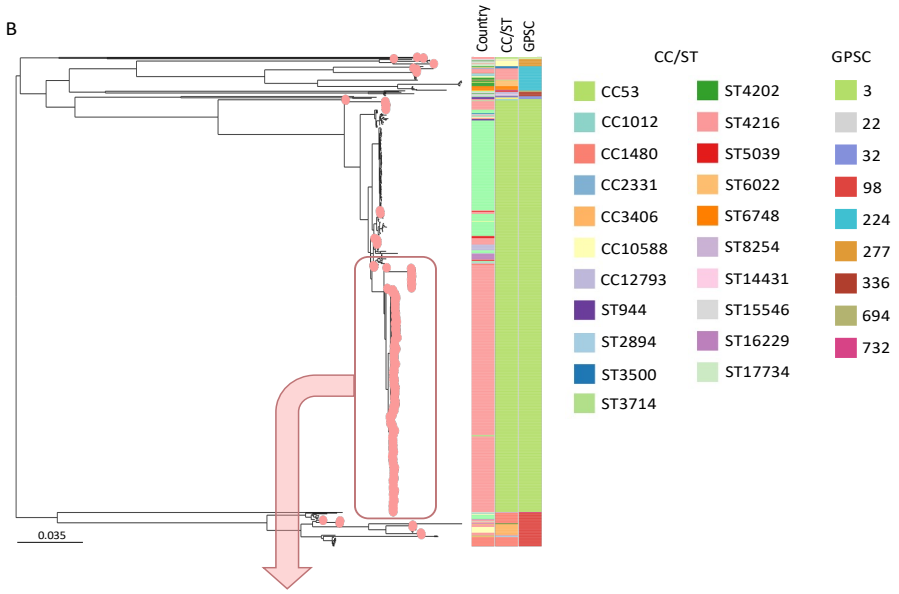


Figure 5.1: Serotype 8 trends and characteristics, 2005–2020, South Africa. a) Incidence of serotype 8 invasive pneumococcal disease reported by year and by age group (N=2429). PCV7 and PCV13 introduction years are shown by dotted lines. Pre-PCV (2005–2008) and late-PCV13 (2015–2020) periods are highlighted in light grey; the COVID-19 pandemic is highlighted in dark grey. b) Known serotype 8 clinical syndromes by age group (n=1365 of 2429). c) Number of sequenced serotype 8 isolates (n=391) by year, stratified by global pneumococcal sequencing clusters (GPSC). PCV7 and PCV13 introduction years are shown by dotted lines.

A


Bangladesh
Belarus
Brazil
Bulgaria
China
Egypt
Ethiopia
France
India
Israel
Malawi
Malaysia
Mongolia
Morocco
Mozambique
Nepal
Netherlands
New Zealand
Nigeria
Papua New Guinea
Poland
Russian Federation
Slovenia
South Africa
Thailand
The Gambia
Turkey
United States

B


Country
CC/ST
GPSC
0.035
CC/ST
CC53
CC1012
CC1480
CC2331
CC3406
CC10588
CC12793
ST944
ST2894
ST3500
ST3714
ST4202
ST4216
ST5039
ST6022
ST6748
ST8254
ST14431
ST15546
ST16229
ST17734
GPSC
3
22
32
98
224
277
336
694
732

C

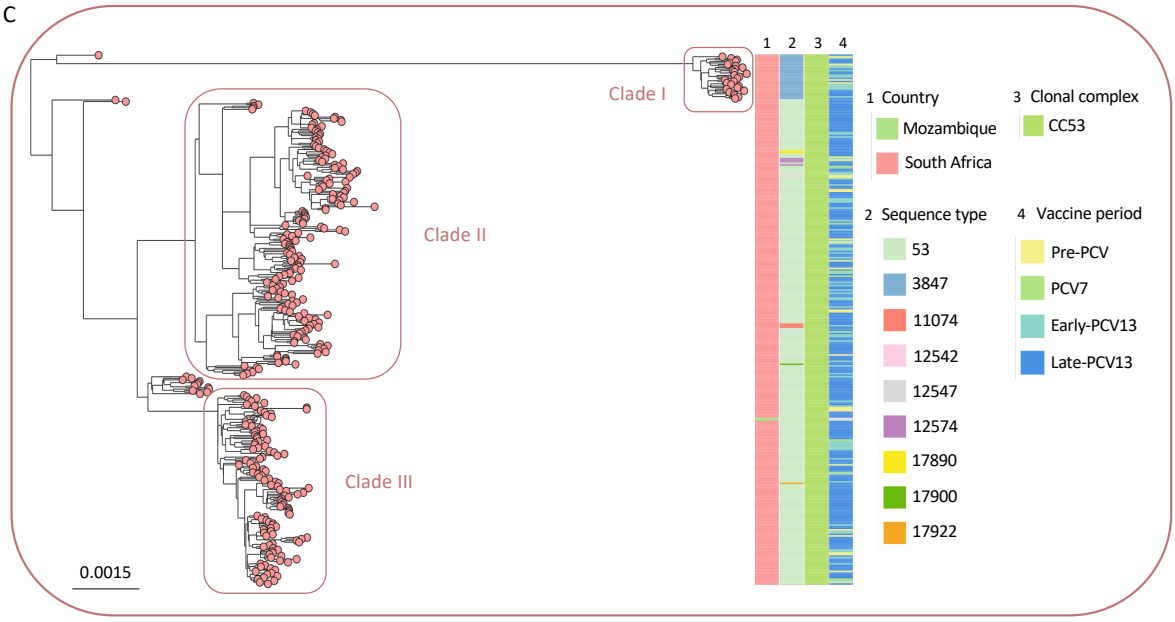

Clade I
Clade II
Clade III
0.0015
1 2 3 4
1 Country
Mozambique
South Africa
2 Sequence type
53
3847
11074
12542
12547
12574
17890
17900
17922
3 Clonal complex
CC53
4 Vaccine period
Pre-PCV
PCV7
Early-PCV13
Late-PCV13

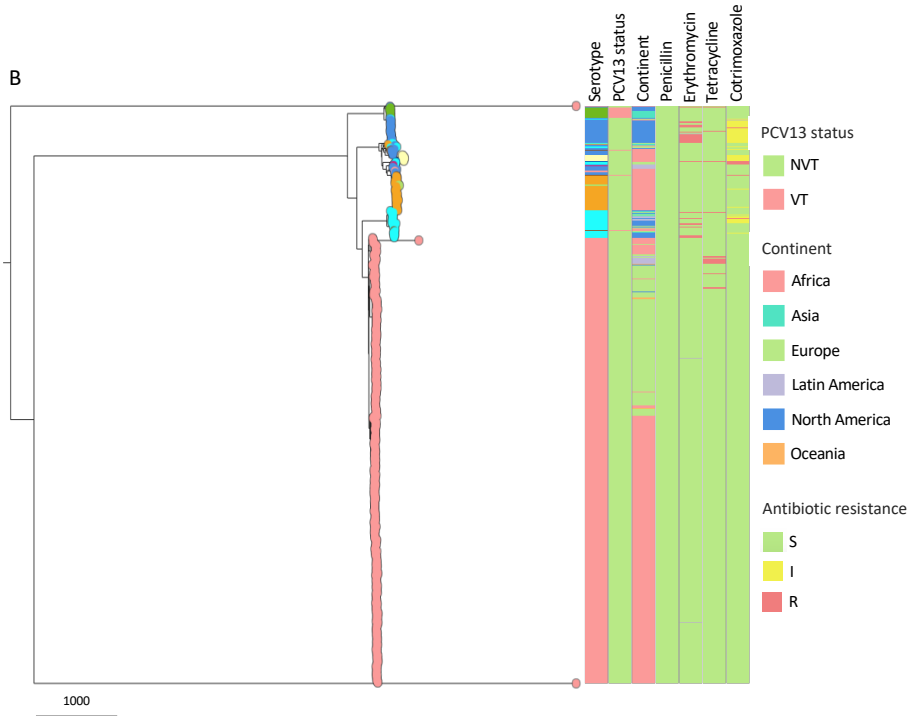
Figure 5.2: Contextualisation of South African serotype 8 genomes. a) Distribution of serotype 8 genomes from South Africa (n=391) and 27 other countries (n=295). The coloured circles on the map are proportional to the number of sequenced isolates by country. b) Phylogeny of all serotype 8 genomes (N=686) with only South African leaf nodes highlighted. c) Expanded South African serotype 8 subtree (n=343) showing country-specific genomic structure. Distinct clades (I through III) are circled and labelled in pink. ST=Sequence type. CC=Clonal complex. GPSC=Global Pneumococcal Sequence Cluster.

*Microreact instance link for phylogeny visualisation: Serotype_8_global_phylogeny

A



B



C

Serotype	Africa	Asia	Europe	Latin America	North America	Oceania
3	3	-	-	-	-	-
8	376	-	188	9	3	5
11A	2	5	13	5	17	-
15A	8	-	-	-	-	-
18C	-	9	1	-	4	-
22F	43	1	-	-	-	-
23F	-	-	-	2	-	-
31	2	-	-	-	-	-
33A	1	-	-	-	-	-
33F	15	1	-	4	32	-
Total	450	16	202	20	56	5

Figure 5.3: Global distribution of GPSC3 lineage (N=749). a) The pie charts are coloured by serotypes expressed by GPSC3 and are proportional to the number of isolates in each country. b) The phylogeny shows the genetic relatedness of GPSC3 serotypes, globally, with metadata inclusive of *in silico* antimicrobial resistance profiles for the antibiotics of interest. c) Shows the distribution of serotypes by continent.

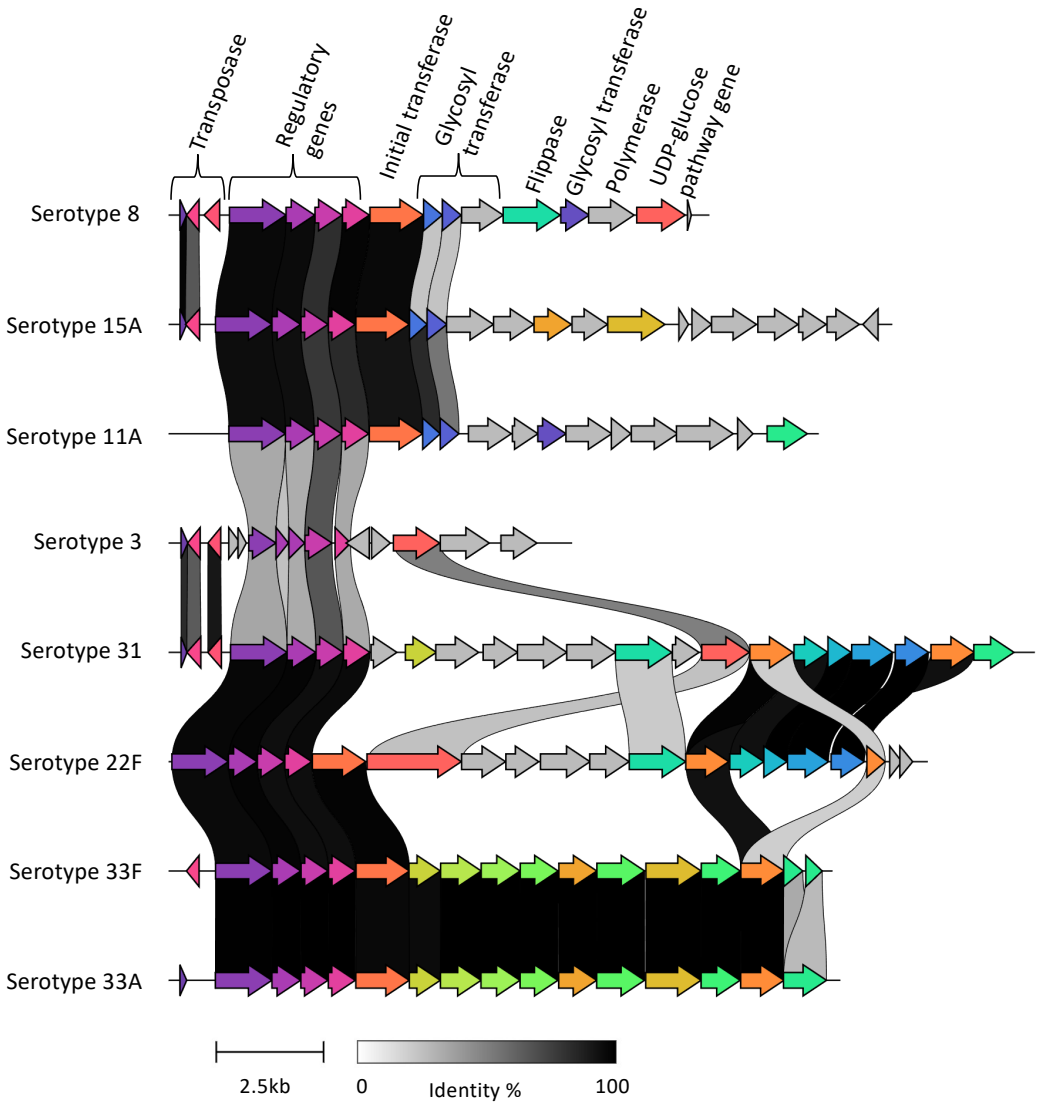


Figure 5.4: GPSC3 *cps* loci comparative analysis. Serotype specific annotations of the eight serotypes expressed by GPSC3 are shown in Appendix V.

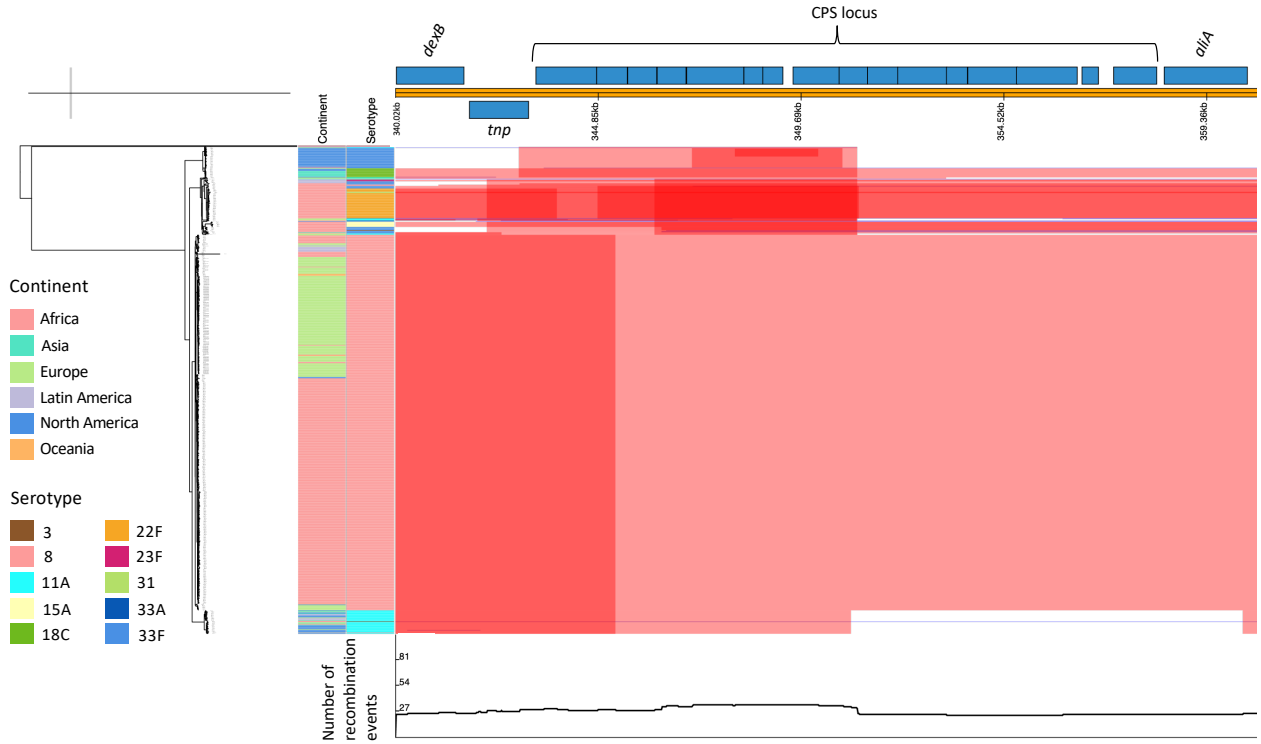


Figure 5.5: GPSC3 Phandango plot of recombination detected with Gubbins, focused on the *cps* locus, across GPSC3 phylogeny, N=749. Phylogenetic tree is on the left, and reference annotation is across the top. Red bars indicate recombinations occurring on internal branches. Blue bars indicate recombinations occurring on terminal branches. The bottom graph shows the number of recombination events.

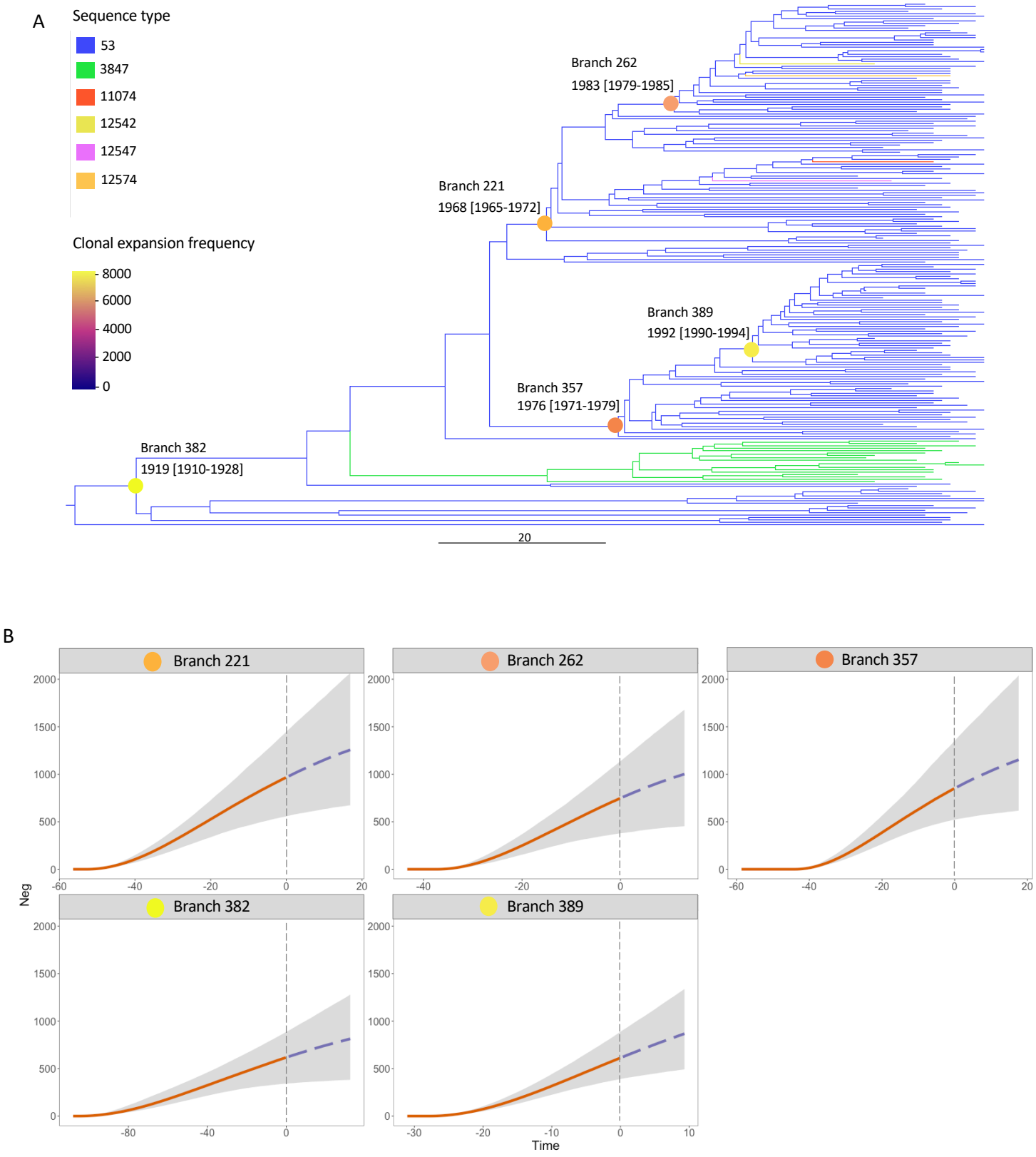


Figure 5.6: CC53 expansion dynamics among genomes from South Africa (N=219) inferred using CaveDive. a) Time-calibrated phylogeny showing clonal expansion branches within

CC53. Clonal expansion frequencies per branch are shown by the colour scale on the left. Phylogeny clades are coloured by sequence type. The phylogenetic tree is shown in increasing node order. b) Expansion dynamics of CC53 subpopulations (expansion branches) showing the effective population sizes over time. Time zero (grey dashed vertical line) represents the year 2020 (end of the study period). The orange solid line represents median past population dynamics, and the purple dashed line represents the predicted population dynamics. The grey shaded area represents 95% credible intervals. Neg=effective population size. CC=clonal complex.

5.4. Discussion

Our study shows that the increase in serotype 8 invasive disease in South Africa after the introduction of PCV was driven by expansion of the pre-existing, globally disseminated GPSC3 lineage. Although GPSC3 largely expressed serotype 8 in our collection, this lineage also expresses other VTs and NVTs, with varying geographical distribution and antimicrobial resistance profiles. Serotype 8 isolates from South Africa showed clonality when contextualised in relation to isolates from other countries, suggesting that expansion was largely due to within-country transmission rather than transmission between countries. Serotype 8 (GPSC3, CC53) remains largely susceptible to commonly used antibiotics globally but continues to be a major cause of invasive disease after PCV introduction, with future predictions showing increases in the effective population size in the absence of serotype 8 targeting interventions.

The increase in NVT disease in both children and adults post PCV vaccination in several countries continues to reduce the net effectiveness of PCVs [58,141,142,191]. The emergence and expansion of serotype 8 (GPSC3, CC53) after PCV introduction in South Africa and other settings demonstrate the continued successful propagation and niche occupation in the PCV era [99,193]. In South Africa, serotype 8 became the most common cause of invasive disease in 2015 (four years after the introduction of PCV13) and accounted for 11% of overall IPD in our surveillance programme. By 2020, serotype 8 IPD proportion had reached 17%. Globally, the proportion of serotype 8 IPD had increased by 120% in 2017 according to The European Centre for Disease Prevention and Control [195].

Serotype 8 is rarely reported in carriage and occasionally associated with outbreaks [196–198]. A meta-analysis of data collected from 20 locations across Africa, Europe, Latin America and North America found that among 25 pneumococcal serotypes, serotype 8 had the 5th highest invasive disease potential estimate after PCV13 serotypes 1, 5, 7F and NVT 12F [199]. We observed different clinical presentations for serotype 8 IPD among children and adults. Studies have shown that serotype 8 is hypervirulent with a thick capsule [200]. This may influence resistance to opsonophagocytosis among children due to immature immune systems resulting in increased likelihood of central nervous system invasion causing meningitis as opposed to blood stream invasion and mucosal persistence in adults. In France, serotype 8 together with emerging non-PCV13 serotypes 12F, 24F, and 33F (also known to have high invasive disease potential) frequently caused bacteraemic pneumonia in patients without underlying conditions [199,201]. In addition, serotype 8 was the most common cause of bacteraemia in Argentinean adults (≥18 years old) in 2013 to 2017 [202].

The GPSC3 (CC53) lineage of serotype 8 is among the most successful lineages in the post-PCV era, globally. Amid global expansion and dissemination, GPSC3 (CC53) isolates from South Africa exhibited distinct clustering when contextualised within a global collection of GPSC3 isolates suggesting within-country transmission. Phylodynamic analysis of CC53 isolates from South Africa showed that this clone had at least five clonal expansions in the past since the most recent common ancestor around 1929. The most recent clonal expansion occurred between 1990 and 1994, with predicted future population dynamics showing continued increase in the effective population sizes of each subpopulation over the next ten years, at least, starting from 2020, implying that some subpopulations may have not reached their carrying capacities.

Although serotype 8 (GPSC3) remains largely susceptible to antimicrobials, it continues to be a major cause of IPD, especially adult IPD in many European countries [141,193,203], and a leading cause of invasive disease among South African children and adults [47,99]. Additionally, the possibility of developing antimicrobial resistance exists. For example, a multidrug-resistant recombinant clone of serotype 8 (ST63), that arose due to recombination between a donor Netherlands 8-ST53 clone and a recipient Sweden 15A-ST63 clone, expanded in Spain after the initial use of PCV in 2001 with prominent spread in 2005, although ST53 maintained predominance during the 2003–2012 period [204]. Similarly, we observed macrolide-nonsusceptibility among GPSC3-expressed NVTs 11A (CC62) and 33F (CC100, ST717 and ST2705) at serotype-specific proportions of 19% and 37%, respectively. GPSC3 has previously been shown to have a lower than average recombination rate (r/m of 5.33 vs median r/m of 7.78, 95% CI 6.81-8.71) as determined from 60 dominant GPSCs using the Wilcoxon rank sum test [105]. However, this lineage does demonstrate the ability to gain or lose genes as shown by the expression of multiple serotypes, both VTs and NVTs.

This study had some limitations. Serotype 8 and GPSC3 incidences were determined using the total number of cases with serotyping data from our GERMS-SA surveillance programme and did not account for those without serotyping data, although missing serotyping data was assumed to have occurred at random. We used a subset of serotype 8 genomes from South Africa to describe circulating lineages and therefore may have missed emerging or rare serotype 8 lineages. From the global collection, serotype 8 and GPSC3 isolates were not systematically selected in some countries which limited our ability to infer lineage predominance compared to South Africa. Nonetheless, our genomic findings were

comparable to those from Spain and Denmark [192,193], identifying serotype 8 GPSC3 (CC53) as the most prevalent disease-causing lineage post-PCV, with other serotype 8 lineages having little disease contribution.

Serotype 8 has substantially contributed to NVT invasive disease in both children and adults in the post-PCV era, globally, while remaining largely susceptible to first-line treatment. The reasons behind this successful propagation are not fully understood. Serotype 8 is targeted by the PCV20 formulation that is recommended for use in the United States and other European countries and because the CPS locus of serotype 8 is largely conserved, chances of vaccine escape may be low. However, South Africa switched to a lower-valency, but affordable PCV10 (Pneumosil, Serum Institute of India) formulation in 2024 which does not include serotype 8. Continued surveillance will be particularly important to monitor the impact of this switch on serotype, lineage and AMR distributions, and may inform the use of higher valency PCVs that target this serotype or the inclusion of serotype 8 in lower valency formulations.

Chapter six - Conclusion

S. pneumoniae remains an important cause of lower respiratory infection, bacteraemia and meningitis despite substantial disease reductions following PCV use (188). A wealth of literature exists on the effectiveness of PCVs against VT disease and epidemiological trend studies that also report on residual VT disease and increases in NVT disease after PCV introduction. Genomic epidemiology studies complementing these findings are widely available from high-income settings and provide insights into the genomic characteristics of clones and lineages that continue to cause disease, some of which benefit from vaccine-selective pressure [90]. In low- and middle-income countries, however, especially in the African continent, studies on genomic surveillance of *S. pneumoniae* before and after PCV introduction are limited. From the earlier, predominantly pre-PCV molecular epidemiology literature from South Africa, it was reported that some of the major disease-causing clones circulating in our setting were novel or rarely reported in other countries [98,100]. We aimed to leverage the high-resolution afforded by next-generation sequencing technologies to evaluate genome-level changes in the pneumococcal population structure in response to PCV introduction in South Africa to better understand post-PCV disease trends which may contribute towards future vaccine design strategies and country-specific vaccine formulation considerations.

Epidemiological studies from South Africa have shown significant declines in VT IPD. This study showed concomitant significant incidence reductions in vaccine-type lineages following the introduction of PCV7 and PCV13. However, after almost a decade post PCV13

use, invasive disease persisted in South Africa and was caused by both pre-existing non-vaccine-type lineages and vaccine-type lineages that additionally expressed NVT variants that shared similar genomic backgrounds with the vacated VT counterparts and possessed similar antibiotic- and vaccine-selection pressure advantages. Examples include the expansion of penicillin-nonsusceptible NVT 35B and MDR serogroup 24 expressed by GPSC5 and GPSC10 lineages that previously predominantly expressed penicillin-nonsusceptible VT 23F and MDR VT 14 in the pre-PCV era, respectively, which may explain their expansion. Endemic penicillin-nonsusceptible or MDR vaccine-type lineages such GPSC17 (VT 19A, CC2062) and GPSC21 (VT 19F, CC347) that circulated before vaccine introduction, continued to cause disease in South Africa in the post-PCV era despite the inclusion of these serotypes in the PCV13 formulation. Future work may include evaluating the genomic features that may be contributing to the persistence of these VTs as similar findings were reported in other countries [157]. The emergence and expansion of penicillin-nonsusceptible or penicillin-MDR serotypes that are not covered by PCV13 were also observed in South Africa and may threaten the benefit of nonsusceptibility reduction as the frequencies of these serotypes may continue to increase with time.

Although antibiotic-nonsusceptible pneumococcal strains contributed notably to IPD in the post-PCV era, non-PCV13 serotype 8 (GPSC3) that was largely susceptible to first-line treatment antibiotics substantially contributed to invasive disease in both children and adults in South Africa after PCV13 introduction. Other settings have also reported on the expansion of this lineage after PCV introduction [192,193]. NVT 8 isolates from South Africa had a clonal genomic population structure when contextualised within a global collection of NVT 8 isolates, suggesting that expansion was largely due to within-country transmission rather than transmission between countries.

Higher valency PCVs such as PCV20 that has been licensed for use in some first-world countries target some of the important serotypes still causing disease in South Africa in the post-PCV era. However, the more affordable SII PCV10 (Serum Institute of India, Pneumosil) that does not include serotypes 3, 4, and 18C replaced PCV13 in 2024. Genomic surveillance will be particularly important in our setting to monitor the impact of switching to a lower valency vaccine, especially since important contributors to overall IPD and nonsusceptible IPD in the post-PCV13 era such NVTs 8, 10A, 15A, 15B/C, 16F, 23B, and 35B are not included in SII PCV10. Post-2020 surveillance reports have shown that NVT 8 remains the most common cause of IPD in children, followed by VTs 3 and 19F. Similarly, NVT 8 and VTs 3 and 4 were common causes of IPD in older children and adults [205,206]. Collaborative genomic epidemiology efforts within the African continent may be useful for in-depth evaluation of endemic lineages and phylogenetics may provide insights into between country transmission dynamics and lineage prevalence. Not only will this effort improve geographic representation, it may also contribute towards vaccine design strategies that prioritize the African continent. Future work may include streamlining a pneumococcus bioinformatics workflow at the NICD for further exploration of pneumococcal genomics in a public health context.

References

1. Euzeby JP. List of bacterial names with standing in nomenclature. Int J Syst Bacteriol. 1997. https://cir.nii.ac.jp/crid/1574231874562909824.

2. Nunes MC, Jones SA, Groome MJ, Kuwanda L, Van Niekerk N, von Gottberg A, et al. Acquisition of Streptococcus pneumoniae in South African children vaccinated with 7-valent pneumococcal conjugate vaccine at 6, 14 and 40 weeks of age. Vaccine. 2015 Jan 29;33(5):628–34.

3. Musher DM, Anderson R, Feldman C. The remarkable history of pneumococcal vaccination: an ongoing challenge. Pneumonia. 2022;14(1).

4. Ganaie F, Saad JS, McGee L, van Tonder AJ, Bentley SD, Lo SW, et al. A new pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral streptococcus. mBio. 2020;11(3).

5. Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. Lancet Infect Dis. 2005;5(2):83–93.

6. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. 2006;2(3):0262–9.

7. Park IH, Kim KH, Andrade AL, Briles DE, Mcdaniel LS, Nahma MH. Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene pspK. mBio. 2012;3(3).

8. Hathaway LJ, Meier PS, Bättig P, Aebi S, Mühlemann K. A homologue of aliB is found in the capsule region of nonencapsulated Streptococcus pneumoniae. J Bacteriol. 2004;186(12):3721–9.

9. Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S, Sá-Leão R, et al. Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol Evol. 2014;6(12):3281–94.

10. Waite RD, Penfold DW, Struthers JK, Dowson CG. Spontaneous sequence duplications within capsule genes cap8E and tts control phase variation in Streptococcus pneumoniae serotypes 8 and 37. Microbiology (Reading). 2003;149(Pt 2):497–504.

11. Sleeman KL, Daniels L, Gardiner M, Griffiths D, Deeks JJ, Dagan R, et al. Acquisition of Streptococcus pneumoniae and nonspecific morbidity in infants and their families: A cohort study. Pediatric Infectious Disease Journal. 2005;24(2):121–7.

12. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect Immun. 2010;78(2):704–15.

13. Bogaert D, De Groot R, Hermans PWM. Streptococcus pneumoniae colonisation: The key to pneumococcal disease. Lancet Infectious Diseases. 2004;4(3):144–54.

14. Regev-Yochay G, Raz M, Dagan R, Porat N, Shainberg B, Pinco E, et al. Nasopharyngeal carriage of Streptococcus pneumoniae by adults and children in community and family settings. Clin Infect Dis. 2004;38(5):632–9.

15. Eng P, Lim LH, Loo CM, Low JA, Tan C, Tan EK, et al. Role of pneumococcal vaccination in prevention of pneumococcal disease among adults in Singapore. Int J Gen Med. 2014;7:179–91.

16. van der Poll T, Opal SM. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. The Lancet. 2009.

17. Paton J, Andrew P, Boulnois G, Mitchell T. Molecular analysis of the pathogenicity of Streptococcus pneumoniae: the role of pneumococcal proteins. Annu Rev Microbiol. 1993.

18. Lynch JP, Zhanel GG. Streptococcus pneumoniae: Epidemiology, risk factors, and strategies for prevention. Semin Respir Crit Care Med. 2009;30(2):189–209.

19. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. The Lancet. 2009;374(9693):893–902.

20. Johnson HL, Deloria-Knoll M, Levine OS, Stoszek SK, Freimanis Hance L, Reithinger R, et al. Systematic Evaluation of Serotypes Causing Invasive Pneumococcal Disease among Children Under Five: The Pneumococcal Global Serotype Project. Cohen J, editor. PLoS Med. 2010;7(10):e1000348.

21. Bryce J, Boschi-Pinto C, Shibuya K, Black RE. WHO estimates of the causes of death in children. Lancet. 2005;365(9465):1147–52.

22. von Gottberg A, Cohen C, De Gouveia L, Meiring S, Quan V, Whitelaw A, et al. Effects of vaccination on invasive pneumococcal disease in South Africa. Vaccine. 2014;371(20):4200–8.

23. Iroh Tam PY, Thielen BK, Obaro SK, Brearley AM, Kaizer AM, Chu H, et al. Childhood pneumococcal disease in Africa – A systematic review and meta-analysis of incidence, serotype distribution, and antimicrobial susceptibility. Vaccine. 2017;35(15):1817–27.

24. Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which Pneumococcal Serogroups Cause the Most Invasive Disease: Implications for Conjugate Vaccine Formulation and Use, Part I. Clinical Infectious Diseases. 2000;30(1):100–21.

25. Johnson HL, Deloria-Knoll M, Levine OS, Stoszek SK, Hance LF, Reithinger R, et al. Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. PLoS medicine. 2010;7(10).

26. Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis. 2010;201(1):32–41.

27. Adegbola RA, Hill PC, Secka O, Ikumapayi UN, Lahai G, Greenwood BM, et al. Serotype and antimicrobial susceptibility patterns of isolates of Streptococcus pneumoniae causing invasive disease in The Gambia 1996-2003. Trop Med Int Health. 2006;11(7):1128–35.

28. Von Mollendorf C, Tempia S, Von Gottberg A, Meiring S, Quan V, Feldman C, et al. Estimated severe pneumococcal disease cases and deaths before and after pneumococcal conjugate vaccine introduction in children younger than 5 years of age in South Africa. PLoS One. 2017;12(7).

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

30. Meiring S, Cohen C, Quan V, De Gouveia L, Feldman C, Karstaedt A, et al. HIV infection and the epidemiology of invasive pneumococcal disease (IPD) in South African adults and

older children prior to the introduction of a pneumococcal conjugate vaccine (PCV). PLoS One. 2016;11(2).

31. Grabenstein JD, Klugman KP. A century of pneumococcal vaccination research in humans. Vol. 18, Clinical Microbiology and Infection. Blackwell Publishing Ltd; 2012. p. 15–24.

32. Mirsaeidi M, Schraufnagel DE. Pneumococcal vaccines: Understanding centers for disease control and prevention recommendations. Ann Am Thorac Soc. 2014;11(6):969–74.

33. Leggat DJ, Khaskhely NM, Iyer AS, Mosakowski J, Thompson RS, Weinandy JD, et al. Pneumococcal polysaccharide vaccination induces polysaccharide-specific B cells in adult peripheral blood expressing CD19+CD20+CD3−CD70−CD27+IgM+CD43+CD5+/−. Vaccine. 2013;31(41):4632–40.

34. Örtqvist Å. Pneumococcal vaccination: current and future issues. Eur Respir J. 2001;18(1):184–95.

35. Fedson DS, Guppy MJ. Pneumococcal vaccination of older adults: Conjugate or polysaccharide? Vol. 9, Human Vaccines and Immunotherapeutics. 2013. p. 1382–4.

36. Feldman C, Dlamini SK, Madhi SA, Meiring S, Von Gottberg A, De Beer JC, et al. The cost-effectiveness of using pneumococcal conjugate vaccine (PCV13) versus pneumococcal polysaccharide vaccine (PPSV23), in South African adults. Vol. 15, PLoS ONE. 2020. 1–49 p.

37. Boyles TH, Brink A, Calligaro GL, Cohen C, Dheda K, Maartens G, et al. South African guideline for the management of community- acquired pneumonia in adults. S Afr J Infect Dis. 2017;9(A Iii):1469–502.

38. Centres for Disease Control and Prevention. Licensure of a 13-valent pneumococcal conjugate vaccine (PCV13) and recommendations for use among children - Advisory Committee on Immunization Practices (ACIP), 2010. Morbidity and Mortality Weekly Report. 2010;59(09).

39. Westerink MAJ, Schroeder HW, Nahm MH. Immune Responses to pneumococcal vaccines in children and adults: Rationale for age-specific vaccination. Vol. 3, Aging and Disease. 2012.

40. Kim TH, Johnstone J, Loeb M. Vaccine herd effect. Scand J Infect Dis. 2011;43(9):683–9.

41. Daniels CC, Rogers PD, Shelton CM. A Review of Pneumococcal Vaccines: Current Polysaccharide Vaccine Recommendations and Future Protein Antigens. The Journal of Pediatric Pharmacology and Therapeutics. 2016;21(1):27–35.

42. Fenoll A, Granizo JJ, Aguilar L, Giménez MJ, Aragoneses-Fenoll L, Hanquet G, et al. Temporal trends of invasive Streptococcus pneumoniae serotypes and antimicrobial resistance patterns in Spain from 1979 to 2007. J Clin Microbiol. 2009;47(4):1012–20.

43. Kobayashi M. Use of 15-valent pneumococcal conjugate vaccine among US children: updated recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep. 2022.

44. Fridh AC, Palmborg A, Ta A, Freigofaite D, Warren S, Perdrizet J. An economic evaluation of pneumococcal conjugate vaccines, PCV20 versus PCV15, for the prevention of pneumococcal disease in the Swedish pediatric population. Hum Vaccin Immunother. 2024;20(1):2400751.

45. Clarke E, Bashorun A, Adigweme I, Hydara MB, Umesi A, Futa A, Ochoge M, Obayemi D, Edem B, Saidy-Jah E, Onwuchekwa C. Immunogenicity and safety of a novel ten-valent pneumococcal conjugate vaccine in healthy infants in The Gambia: a phase 3, randomised, double-blind, non-inferiority trial. The Lancet Infectious Diseases. 2021;21(6):834-46.

46. Madhi SA, Cohen C, von Gottberg A. Introduction of pneumococcal conjugate vaccine into the public immunization program in South Africa: Translating research into policy. Vol. 30, Vaccine. Elsevier; 2012. p. C21–7.

47. von Gottberg A, Kleynhans J, de Gouveia L, Tempia S, Meiring S, Quan V, et al. Long-term effect of pneumococcal conjugate vaccines on invasive pneumococcal disease incidence among people of all ages from national, active, laboratory-based surveillance in South Africa, 2005-19: a cohort observational study. Lancet Glob Health. 2024;12(9):e1470–84.

48. Moore MR, Link-Gelles R, Schaffner W, Lynfield R, Lexau C, Bennett NM, et al. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: Analysis of multisite, population-based surveillance. Lancet Infect Dis. 2015;15(3):301–9.

49. Mackenzie GA, Hill PC, Jeffries DJ, Ndiaye M, Sahito SM, Hossain I, et al. Impact of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease and

pneumonia in The Gambia: 10 years of population-based surveillance. Lancet Infect Dis. 2021;21(9):1293–302.

50. Miller E, Andrews NJ, Waight PA, Slack MPE, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: An observational cohort study. Lancet Infect Dis. 2011;11(10):760–8.

51. Chapman R, Sutton K, Dillon-Murphy D, Patel S, Hilton B, Farkouh R, et al. Ten year public health impact of 13-valent pneumococcal conjugate vaccination in infants: A modelling analysis. Vaccine. 2020;38(45):7138–45.

52. Huebner RE, Klugman KP, Matai U, Eggers R, Hussey G, Brink A, et al. Laboratory surveillance for Haemophilus influenzae type B, meningococcal, and pneumococcal disease. Vol. 89, South African Medical Journal. 1999. p. 924–5.

53. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. The Lancet. 2011;378(9807):1962–73.

54. McElligott M, Vickers I, Meehan M, Cafferkey M, Cunney R, Humphreys H. Noninvasive pneumococcal clones associated with antimicrobial nonsusceptibility isolated from children in the era of conjugate vaccines. Antimicrob Agents Chemother. 2015;59(9):5761–7.

55. Kaur R, Casey JR, Pichichero ME. Emerging streptococcus pneumoniae strains colonizing the nasopharynx in children after 13-valent pneumococcal conjugate vaccination in comparison to the 7-valent era, 2006-2015. Pediatric Infectious Disease Journal. 2016;35(8):901–6.

56. Kawaguchiya M, Urushibara N, Kobayashi N. Multidrug Resistance in Non-PCV13 Serotypes of Streptococcus pneumoniae in Northern Japan, 2014. Microbial Drug Resistance. 2017;23(2):206–14.

57. Azarian T, Grant LR, Arnold BJ, Hammitt LL, Reid R, Santosham M, et al. The impact of serotype-specific vaccination on phylodynamic parameters of Streptococcus pneumoniae and the pneumococcal pan-genome. PLoS Pathog. 2018;14(4).

58. Lo SW, Gladstone RA, van Tonder AJ, Lees JA, du Plessis M, Benisty R, et al. Pneumococcal lineages associated with serotype replacement and antibiotic resistance in childhood invasive pneumococcal disease in the post-PCV13 era: an international whole-genome sequencing study. Lancet Infect Dis. 2019;19(7):759–69.

59. Chochua S, Metcalf BJ, Li Z, Walker H, Tran T, McGee L, et al. Invasive serotype 35B pneumococci including an expanding serotype switch lineage, United States, 2015–2016. Emerg Infect Dis. 2017;23(6):922–30.

60. Candeias C, Félix S, Handem S, de Lencastre H, Sá-Leão R. Clonal changes in the pneumococcal population carried by Portuguese children during six years of private use of the 13-valent pneumococcal conjugate vaccine: The relative contribution of clonal expansion, clonal emergence, and capsular switch events. Microbiology Spectrum. 2023;11(2):e02909-22.

61. Pimenta F, Moiane B, Gertz RE, Chochua S, Snippes Vagnone PM, Lynfield R, et al. New Pneumococcal Serotype 15D. J Clin Microbiol. 2021;59(5).

62. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine Escape Recombinants Emerge after Pneumococcal Vaccination in the United States. PLoS Pathog. 2007;3(11):e168.

63. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, Von Gottberg A, Liñares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis. 2013;207(3):439–49.

64. Ampofo K, Pavia AT, Chris S, Hersh AL, Bender JM, Blaschke AJ, et al. The changing epidemiology of invasive pneumococcal disease at a tertiary children's hospital through the 7-valent pneumococcal conjugate vaccine era: A case for continuous surveillance. Pediatric Infectious Disease Journal. 2012;31(3):228–34.

65. Hanquet G, Kissling E, Fenoll A, George R, Lepoutre A, Lernout T, et al. Pediatric Pneumococcal Serotypes in 4 European Countries. Emerg Infect Dis. 2010;16(9):1428.

66. Cho EY, Kang HM, Lee J, Kang JH, Choi EH, Lee HJ. Changes in serotype distribution and antibiotic resistance of nasopharyngeal isolates of Streptococcus pneumoniae from children in Korea, after optional use of the 7-valent conjugate vaccine. J Korean Med Sci. 2012;27(7):716–22.

67. Lehmann D, Willis J, Moore HC, Giele C, Murphy D, Keil AD, et al. The changing epidemiology of invasive pneumococcal disease in aboriginal and non-aboriginal western Australians from 1997 through 2007 and emergence of nonvaccine serotypes. Clin Infect Dis. 2010;50(11):1477–86.

68. Porat N, Benisty R, Trefler R, Givon-Lavi N, Dagan R. Clonal distribution of common pneumococcal serotypes not included in the 7-valent conjugate vaccine (PCV7): marked

differences between two ethnic populations in southern Israel. J Clin Microbiol. 2012;50(11):3472–7.

69. Moore MR, Gertz RE, Woodbury RL, Barkocy-Gallagher GA, Schaffner W, Lexau C, et al. Population snapshot of emergent Streptococcus pneumoniae serotype 19A in the United States, 2005. Journal of Infectious Diseases. 2008;197(7):1016–27.

70. Metcalf BJ, Gertz RE, Gladstone RA, Walker H, Sherwood LK, Jackson D, et al. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. Clinical Microbiology and Infection. 2016;22(1):60.e9-60.e29.

71. Olarte L, Kaplan SL, Barson WJ, Romero JR, Lin PL, Tan TQ, et al. Emergence of multidrug-resistant pneumococcal serotype 35B among children in the United States. J Clin Microbiol. 2017;55(3):724–34.

72. Boyles TH, Bamford C, Bateman K, Blumberg L, Dramowski A, Korsman S, et al. Guidelines for the management of acute meningitis in children and adults in South Africa. 2015;8782.

73. Cornick JE, Bentley SD. Streptococcus pneumoniae: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. Microbes Infect. 2012;14(7–8):573–83.

74. The Center for Disease Dynamics E& P. ResistanceMap: Antibiotic use. 2018. Available from: https://resistancemap.onehealthtrust.org/About.php

75. Hansman D, Bullen MM. A resistant pneumococcus. The Lancet. 1967;2:264–5.

76. Appelbaum PC, Scragg JN, Bowen AJ, Bhamjee A, Hallett AF, Cooper RC. Streptococcus pneumoniae resistant to penicillin and chloramphenicol. The Lancet. 1977;310(8046):995–7.

77. Guillemot D, Carbon C. Antibiotic use and pneumococcal resistance to penicillin: the French experience. Clinical microbiology and infection. 1999;5:4S38-42.

78. Marton A, Gulyas M, Munoz R, Tomasz A. Extremely High Incidence of Antibiotic Resistance in Clinical Isolates of Streptococcus pneumoniae in Hungary. J Infect Dis. 1991;163(3):542–8.

79. Kristinsson KG. Epidemiology of Penicillin Resistant Pneumococci in Iceland. Microbial Drug Resistance. 2009;1(2):121–5.

80. Friedland IR, Klugman KP. Antibiotic-Resistant Pneumococcal Disease in South African Children. American Journal of Diseases of Children. 1992;146(8):920–3.

81. Fenoll A, Jado I, Vicioso D, Pérez A, Casal J. Evolution of Streptococcus pneumoniae serotypes and antibiotic resistance in Spain: update (1990 to 1996). J Clin Microbiol. 1998;36(12):3447–54.

82. Breiman RF, Butler JC, Tenover FC, Elliott JA, Facklam RR. Emergence of Drug-Resistant Pneumococcal Infections in the United States. JAMA. 1994;271(23):1831–5.

83. Song JH, Lee NY, Ichiyama S, Yoshida R, Hirakata Y, Fu W, et al. Spread of Drug-Resistant Streptococcus pneumoniae in Asian Countries: Asian Network for Surveillance of Resistant Pathogens (ANSORP) Study. Clinical Infectious Diseases. 1999;28(6):1206–11.

84. Richter SS, Diekema DJ, Heilmann KP, Dohrn CL, Riahi F, Doern G V. Changes in pneumococcal serotypes and antimicrobial resistance after introduction of the 13-Valent conjugate vaccine in the United States. Antimicrob Agents Chemother. 2014;58(11):6484–9.

85. Bajema KL, Gierke R, Farley MM, Schaffner W, Thomas A, Reingold AL, et al. Impact of Pneumococcal Conjugate Vaccines on Antibiotic-Nonsusceptible Invasive Pneumococcal Disease in the United States. J Infect Dis. 2022;226(2):342–51.

86. Manzanal A, Vicente D, Alonso M, Azkue N, Ercibengoa M, Marimón JM. Impact of the progressive uptake of pneumococcal conjugate vaccines on the epidemiology and antimicrobial resistance of invasive pneumococcal disease in Gipuzkoa, northern Spain, 1998–2022. Front Public Health. 2023;11:1238502.

87. Lo SW, Mellor K, Cohen R, Alonso AR, Belman S, Kumar N, et al. Emergence of a multidrug resistant and virulent Streptococcus pneumoniae lineage mediates serotype replacement after PCV13. medRxiv. 2021;5247(22):2021.11.24.21266813.

88. Beall B, Mc Ellistrem MC, Gertz RE, Boxrud DJ, Besser JM, Harrison LH, et al. Emergence of a Novel Penicillin-Nonsusceptible, Invasive Serotype 35B Clone of Streptococcus pneumoniae within the United States. J Infect Dis. 2002;186(1):118–22.

89. Tomczyk S, Lynfield R, Schaffner W, Reingold A, Miller L, Petit S, et al. Prevention of Antibiotic-Nonsusceptible Invasive Pneumococcal Disease With the 13-Valent Pneumococcal Conjugate Vaccine. Clinical Infectious Diseases. 2016;62(9):1119–25.

90. Obolski U, Lourenço J, Thompson C, Thompson R, Gori A, Gupta S. Vaccination can drive an increase in frequencies of antibiotic resistance among nonvaccine serotypes of Streptococcus pneumoniae. Proc Natl Acad Sci USA. 2018;115(12):3102–7.

91. Reynolds CA, Finkelstein JA, Ray GT, Moore MR, Huang SS. Attributable healthcare utilization and cost of pneumonia due to drug-resistant streptococcus pneumonia: a cost analysis. Antimicrob Resist Infect Control. 2014;3(1).

92. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA. 1998;95(6):3140–5.

93. Dunne EM, Ong EK, Moser RJ, Siba PM, Phuanukoonnon S, Greenhill AR, et al. Multilocus sequence typing of Streptococcus pneumoniae by use of mass spectrometry. J Clin Microbiol. 2011;49(11):3756–60.

94. Pérez-Losada M, Arenas M, Castro-Nallar E. Multilocus Sequence Typing of Pathogens: Methods, Analyses, and Applications. Methods, Analyses, and Applications. Genetics and Evolution of Infectious Diseases: Second Edition. 2017;383–404.

95. Spratt BG, Hanage WP, Li B, Aanensen DM, Feil EJ. Displaying the relatedness among isolates of bacterial species–the eBURST approach. FEMS microbiology letters. 2004;241(2):129-34.

96. Scott JR, Hanage WP, Lipsitch M, Millar E V., Moulton LH, Hinds J, et al. Pneumococcal sequence type replacement among American Indian children: A comparison of pre- and routine-PCV7 eras. Vaccine. 2012;30(13):2376–81.

97. Beall BW, Gertz RE, Hulkower RL, Whitney CG, Moore MR, Brueggemann AB. Shifting Genetic Structure of Invasive Serotype 19A Pneumococci in the United States. J Infect Dis. 2011;203(10):1360–8.

98. Ndlangisa KM, du Plessis M, Lo S, de Gouveia L, Chaguza C, Antonio M, et al. A Streptococcus pneumoniae lineage usually associated with pneumococcal conjugate vaccine (PCV) serotypes is the most common cause of serotype 35B invasive disease in South Africa, following routine use of PCV. Microb Genom. 2022;8(4):1–10.

99. Lekhuleni C, Ndlangisa K, Gladstone RA, Chochua S, Metcalf BJ, Li Y, et al. Impact of pneumococcal conjugate vaccines on invasive pneumococcal disease-causing lineages among South African children. Nat Commun. 2024;15(1):1–12.

100. Ndlangisa KM, Du Plessis M, Wolter N, De Gouveia L, Klugman KP, Von Gottberg A. Population snapshot of streptococcus pneumoniae causing invasive disease in South Africa prior to introduction of pneumococcal conjugate vaccines. PLoS One. 2014;9(9):1–11.

101. Mothibeli KM, du Plessis M, von Gottberg A, de Gouveia L, Adrian P, Madhi SA, et al. An Unusual Pneumococcal Sequence Type Is the Predominant Cause of Serotype 3 Invasive Disease in South Africa. J Clin Microbiol. 2010;48(1):184–91.

102. Du Plessis M, Allam M, Tempia S, Wolter N, De Gouveia L, Von Mollendorf C, et al. Phylogenetic analysis of invasive serotype 1 pneumococcus in South Africa, 1989 to 2013. J Clin Microbiol. 2016;54(5):1326–34.

103. Yaro S, Lourd M, Traoré Y, Njanpop-Lafourcade BM, Sawadogo A, Sangare L, et al. Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso. Clinical Infectious Diseases. 2006;43(6):693–700.

104. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, Van Der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science (1979). 2011;331(6016):430–4.

105. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. EBioMedicine. 2019;43:338–46.

106. Pearce ME, Alikhan NF, Dallman TJ, Zhou Z, Grant K, Maiden MCJ. Comparative analysis of core genome MLST and SNP typing within a European Salmonella serovar Enteritidis outbreak. Int J Food Microbiol. 2018;274:1–11.

107. Ashton PM, Nair S, Peters TM, Bale JA, Powell DG, Painset A, et al. Identification of Salmonella for public health surveillance using whole genome sequencing. PeerJ. 2016;2016(4):e1752.

108. Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, Boxrud D. Characterization of foodborne outbreaks of Salmonella enterica serovar Enteritidis with

whole-genome sequencing single nucleotide polymorphism-based analysis for surveillance and outbreak detection. Journal of clinical microbiology. 2015;53(10):3334-40.

109. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome research. 2019;29(2):304-16.

110. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Molecular biology and evolution. 2013;30(5):1224-8.

111. Hawkins PA, Chochua S, Lo SW, Belman S, Antonio M, Kwambana-Adams B, et al. A global genomic perspective on the multidrug-resistant Streptococcus pneumoniae 15A-CC63 sub-lineage following pneumococcal conjugate vaccine introduction. Microb Genom. 2023;9(4).

112. Cave R, Kalizang'oma A, Chaguza C, Mwalukomo TS, Kamng'ona A, Brown C, et al. Expansion of pneumococcal serotype 23F and 14 lineages with genotypic changes in capsule polysaccharide locus and virulence gene profiles post introduction of pneumococcal conjugate vaccine in Blantyre, Malawi. Microb Genom. 2024;10(6).

113. Rückinger S, von Kries R, Reinert RR, van der Linden M, Siedler A. Childhood invasive pneumococcal disease in Germany between 1997 and 2003: Variability in incidence and serotype distribution in absence of general pneumococcal conjugate vaccination. Vaccine. 2008;26(32):3984–6.

114. Henriques-Normark B, Blomberg C, Dagerhamn J, Bättig P, Normark S. The rise and fall of bacterial clones: Streptococcus pneumoniae. Nat Rev Microbiol. 2008;6(11):827–37.

115. Jefferies JM, Smith AJ, Edwards GFS, McMenamin J, Mitchell TJ, Clarke SC. Temporal analysis of invasive pneumococcal clones from Scotland illustrates fluctuations in diversity of serotype and genotype in the absence of pneumococcal conjugate vaccine. J Clin Microbiol. 2010;48(1):87–96.

116. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations. Mol Biol Evol. 2017;34(5):1167–82.

117. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43(3):e15.

118. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):1–16.

119. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58(1):212–20.

120. Epping L, van Tonder AJ, Gladstone RA, Bentley SD, Page AJ, Keane JA. SeroBA: Rapid high-throughput serotyping of Streptococcus pneumoniae from whole genome sequence data. Microb Genom. 2018;4(7):1–6.

121. Page AJ, Taylor B, Keane JA. Multilocus sequence typing by blast from de novo assemblies against PubMLST. The Journal of Open Source Software. 2016;1(8):118.

122. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3(10).

123. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. 2018;46(22).

124. Volz EM, Didelot X. Modeling the Growth and Decline of Pathogen Effective Population Size Provides Insight into Epidemic Dynamics and Drivers of Antimicrobial Resistance. Syst Biol. 2018;67(4):719–28.

125. Helekal D, Ledda A, Volz E, Wyllie D, Didelot X. Bayesian Inference of Clonal Expansions in a Dated Phylogeny. Syst Biol. 2022;71(5):1073.

126. Carvalho MDGS, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, et al. Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. J Clin Microbiol. 2007;45(8):2460–6.

127. Austrian R. The Quellung reaction, a neglected microbiologic technique. Mount Sinai Journal of Medicine. 1976;43(6):699–709.

128. Azzari C, Moriondo M, Indolfi G, Cortimiglia M, Canessa C, Becciolini L, et al. Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. PLoS One. 2010;5(2).

129. Pimenta FC, Roundtree A, Soysal A, Bakir M, Du Plessis M, Wolter N, et al. Sequential Triplex Real-Time PCR Assay for Detecting 21 Pneumococcal Capsular Serotypes That Account for a High Global Disease Burden. J Clin Microbiol. 2013;51(2):647–52.

130. Pholwat S, Sakai F, Turner P, Vidal JE, Houpt ER. Development of a taqman array card for pneumococcal serotyping on isolates and nasopharyngeal samples. J Clin Microbiol. 2016;54(7):1842–50.

131. Clinical and Laboratory Standards Institute. Performance Standards for Antimicrobial Susceptibility Testing: 28th Edition Informational Supplement M100-S27; CLSI: Wayne, PA, USA, 2018. Clinical and Laboratory Standards Institute. 2018.

132. Zerbino DR. Using the Velvet de novo Assembler for Short-Read Sequencing Technologies. Curr Protoc Bioinformatics. 2010;31(1):11.5.1-11.5.12.

133. Gladman S, Seemann T. http://bioinformatics.net.au/software.velvetoptimiser.shtml.

134. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27(4):578-9.

135. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012;13(6).

136. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.

137. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic acids research. 2012;40(D1):D130-5.

138. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. J Bacteriol. 2004;186(5):1518–30.

139. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Penicillin-binding protein transpeptidase signatures for tracking and predicting β-lactam resistance levels in Streptococcus pneumoniae. mBio. 2016;7(3):1–9.

140. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Validation of β-lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. BMC Genomics. 2017;18(1):621.

141. Ladhani SN, Collins S, Djennad A, Sheppard CL, Borrow R, Fry NK, et al. Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000–17: a prospective national observational cohort study. Lancet Infect Dis. 2018;18(4):441–51.

142. Mackenzie GA, Hill PC, Jeffries DJ, Hossain I, Uchendu U, Ameh D, et al. Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study. Lancet Infect Dis. 2016;16(6):703–11.

143. Corander J, Fraser C, Gutmann MU, Arnold B, Hanage WP, Bentley SD, et al. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. Nat Ecol Evol. 2017;1(12):1950–60.

144. Ho SYW, Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide sequences. Vol. 11, Molecular Ecology Resources. 2011. p. 423–34.

145. Golden AR, Baxter MR, Davidson RJ, Martin I, Demczuk W, Mulvey MR, Karlowsky JA, Hoban DJ, Zhanel GG, Adam HJ. Comparison of antimicrobial resistance patterns in Streptococcus pneumoniae from respiratory and blood cultures in Canadian hospitals from 2007–16. Journal of Antimicrobial Chemotherapy. 2019;74(Supplement_4):iv39-47.

146. Kaur R, Pham M, Yu KOA, Pichichero ME. Rising Pneumococcal Antibiotic Resistance in the Post-13-Valent Pneumococcal Conjugate Vaccine Era in Pediatric Isolates From a Primary Care Setting. Clin Infect Dis. 2021;72(5):797–805.

147. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom. 2016;2(4):e000056.

148. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3).

149. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016;2(11):e000093.

150. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

151. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.

152. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics. 2018; 15;34(2):292–3.

153. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R news. 2006;6(1):7-11.

154. Cohen C, von Mollendorf C, de Gouveia L, Lengana S, Meiring S, Quan V, et al. Effectiveness of the 13-valent pneumococcal conjugate vaccine against invasive pneumococcal disease in South African children: a case-control study. Lancet Glob Health. 2017;5(3):e359–69.

155. Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, Lynfield R, et al. Decline in Invasive Pneumococcal Disease after the Introduction of Protein–Polysaccharide Conjugate Vaccine. New England Journal of Medicine. 2003;348(18):1737–46.

156. Andrews NJ, Waight PA, Burbidge P, Pearce E, Roalfe L, Zancolli M, et al. Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: A postlicensure indirect cohort study. Lancet Infect Dis. 2014;14(9):839–46.

157. Corcoran M, Mereckiene J, Cotter S, Murchan S, Lo SW, McGee L, et al. Using genomics to examine the persistence of Streptococcus pneumoniae serotype 19A in Ireland and the emergence of a sub-clade associated with vaccine failures. Vaccine. 2021;39(35):5064–73.

158. Pillai DR, Shahinas D, Buzina A, Pollock RA, Lau R, Khairnar K, et al. Genome-wide dissection of globally emergent multi-drug resistant serotype 19A Streptococcus pneumoniae. BMC Genomics. 2009;10:1–13.

159. Hanage WP, Bishop CJ, Lee GM, Lipsitch M, Stevenson A, Rifas-Shiman SL, et al. Clonal replacement among 19A Streptococcus pneumoniae in Massachusetts, prior to 13 valent conjugate vaccination. Vaccine. 2011;29(48):8877–81.

160. Shin J, Baek JY, Kim SH, Song JH, Ko KS. Predominance of ST320 among Streptococcus pneumoniae serotype 19A isolates from 10 Asian countries. Journal of Antimicrobial Chemotherapy. 2011;66(5):1001–4.

161. Hanage WP, Finkelstein JA, Huang SS, Pelton SI, Stevenson AE, Kleinman K, et al. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. Epidemics. 2010;2(2):80–4.

162. Müller A, Kleynhans J, de Gouveia L, Meiring S, Cohen C, Hathaway LJ, et al. Streptococcus pneumoniae Serotypes Associated with Death, South Africa, 2012-2018. Emerg Infect Dis. 2022;28(1):166–79.

163. Croucher NJ, Chewapreecha C, Hanage WP, Harris SR, McGee L, Van Der Linden M, et al. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. Genome Biol Evol. 2014;6(7):1589–602.

164. Kandasamy R, Lo S, Gurung M, Carter MJ, Gladstone R, Lees J, et al. Effect of childhood vaccination and antibiotic use on pneumococcal populations and genome-wide associations with disease among children in Nepal: an observational study. Lancet Microbe. 2022;3(7):e503–11.

165. Wolter N, von Gottberg A, du Plessis M, de Gouveia L, Klugman KP. Molecular basis and clonal nature of increasing pneumococcal macrolide resistance in South Africa, 2000-2005. Int J Antimicrob Agents. 2008;32(1):62–7.

166. Soeters HM, Von Gottberg A, Cohen C, Quan V, Klugman KP. Trimethoprim-sulfamethoxazole prophylaxis and antibiotic nonsusceptibility in invasive pneumococcal disease. Antimicrob Agents Chemother. 2012;56(3):1602–5.

167. The Lancet. Antimicrobial resistance: an agenda for all. The Lancet. 2024;403(10442):2349.

168. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance. 2016. p. 84.

169. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022;399(10325):629–55.

170. Belman S, Lefrancq N, Nzenze S, et al. Geographical migration and fitness dynamics of Streptococcus pneumoniae. Nature. 2024.

171. WHO Bacterial Priority Pathogens List, 2024: bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Geneva; 2024.

172. Jansen KU, Anderson AS. The role of vaccines in fighting antimicrobial resistance (AMR). Hum Vaccin Immunother. 2018;14(9):2142.

173. Micoli F, Bagnoli F, Rappuoli R, Serruto D. The role of vaccines in combatting antimicrobial resistance. Nat Rev Microbiol. 2021;19(5):287–302.

174. Klugman KP. The successful clone: The vector of dissemination of resistance in Streptococcus pneumoniae. Vol. 50, Journal of Antimicrobial Chemotherapy. 2002. p. 1–5.

175. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, et al. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. J Clin Microbiol. 2001;39(7):2565–71.

176. Whitney CG, Farley MM, Hadler J, Harrison LH, Lexau C, Reingold A, et al. Increasing prevalence of multidrug-resistant Streptococcus pneumoniae in the United States. N Engl J Med. 2000;343(26):1917–24.

177. Ho PL, Law PY, Chiu SS. Increase in incidence of invasive pneumococcal disease caused by serotype 3 in children eight years after the introduction of the pneumococcal conjugate vaccine in Hong Kong. Human Vaccines & Immunotherapeutics. 2019;15(2):455-8.

178. Berger Y, Adler A, Ariel T, Rokney A, Averbuch D, Grisaru-Soen G. Paediatric community-acquired bacteraemia, pneumococcal invasive disease and antibiotic resistance fell after the pneumococcal conjugate vaccine was introduced. Acta Paediatr. 2019;108(7):1321–8.

179. Quirk SJ, Haraldsson G, Hjálmarsdóttir MA, Van Tonder AJ, Hrafnkelsson B, Bentley SD, et al. Vaccination of icelandic children with the 10-valent pneumococcal vaccine leads to a significant herd effect among adults in Iceland. J Clin Microbiol. 2018;57(4).

180. Cassiolato AP, Grassi Almeida SC, Andrade AL, Minamisava R, Cristina de Cunto Brandileone M. Expansion of the multidrug-resistant clonal complex 320 among invasive Streptococcus pneumoniae serotype 19A after the introduction of a ten-valent pneumococcal conjugate vaccine in Brazil. PLoS One. 2018;13(11).

181. Koutouzis EI, Daikos GL, Chatzichristou P, Michos AG, Koutouzi FI, Chrousos GP, et al. Characteristics of Streptococcus pneumoniae serotype 19A isolates from children in the pre

and post Conjugate Vaccine Era. Single center experience 1986–2015. Vaccine. 2018;36(35):5245–50.

182. Candeias C, Almeida ST, Paulo AC, Simões AS, Ferreira B, Cruz AR, et al. Streptococcus pneumoniae carriage, serotypes, genotypes, and antimicrobial resistance trends among children in Portugal, after introduction of PCV13 in National Immunization Program: A cross-sectional study. Vaccine. 2024;42(22):126219.

183. Tiley KS, Ratcliffe H, Voysey M, Jefferies K, Sinclair G, Carr M, et al. Nasopharyngeal Carriage of Pneumococcus in Children in England up to 10 Years After 13-Valent Pneumococcal Conjugate Vaccine Introduction: Persistence of Serotypes 3 and 19A and Emergence of 7C. J Infect Dis. 2023;227(5):610–21.

184. Kobayashi M. Use of 15-valent pneumococcal conjugate vaccine and 20-valent pneumococcal conjugate vaccine among US adults: updated recommendations of the Advisory Committee on Immunization Practices—United States, 2022. MMWR. Morbidity and mortality weekly report. 2022;71.

185. Patel SM, Shaik-Dasthagirisaheb YB, Congdon M, Young RR, Patel MZ, Mazhani T, et al. Evolution of pneumococcal serotype epidemiology in Botswana following introduction of 13-valent pneumococcal conjugate vaccine. PLoS One. 2022;17(1).

186. van der Linden M, Perniciaro S, Imöhl M. Increase of serotypes 15A and 23B in IPD in Germany in the PCV13 vaccination era. BMC Infect Dis. 2015;15(1).

187. Gladstone RA, Lo SW, Goater R, Yeats C, Taylor B, Hadfield J, et al. Visualizing variation within global pneumococcal sequence clusters (GPSCs) and country population snapshots to contextualize pneumococcal isolates. Microb Genom. 2020;6(5):1–13.

188. Mokaya J, Mellor KC, Murray GGR, et al. Genomic epidemiology of Streptococcus pneumoniae serotype 16F lineages. Microb Genom. 2023; 9(11):1–9.

189. Daniels B, Kuhn L, Spooner E, et al. Cotrimoxazole guidelines for infants who are HIV-exposed but uninfected: a call for a public health and ethics approach to the evidence. Lancet Glob Health. 2022.

190. D'Aeth JC, Bertran M, Abdullahi F, et al. Whole-genome sequencing, strain composition, and predicted antimicrobial resistance of Streptococcus pneumoniae causing invasive disease in England in 2017–20: a prospective national surveillance study. Lancet Microbe. 2025.

191. Løchen A, Croucher NJ, Anderson RM. Divergent serotype replacement trends and increasing diversity in pneumococcal disease in high income settings reduce the benefit of expanding vaccine valency. Sci Rep. 2020;10(1).

192. Sanz JC, Rodríguez-Avial I, Ríos E, García-Comas L, Ordobás M, Cercenado E. Increase of serotype 8, ST53 clone, as the prevalent strain of Streptococcus pneumoniae causing invasive disease in Madrid, Spain (2012–2015). Enferm Infecc Microbiol Clin. 2020;38(3):105–10.

193. Hansen, C. B., Fuursted, K., Valentiner-Branth, P., Dalby, T., Jørgensen, C. S., & Slotved HC. Molecular characterization and epidemiology of Streptococcus pneumoniae serotype 8 in Denmark. BMC Infect Dis. 2021;21(1).

194. Gilchrist CLM, Chooi YH. clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics. 2021.

195. ECDC. European centre for disease prevention and control. Invasive pneumococcal disease. Annual epidemiological report for 2017. 2019.

196. Lindstrand A, Galanis I, Darenberg J, Morfeldt E, Naucler P, Blennow M, et al. Unaltered pneumococcal carriage prevalence due to expansion of non-vaccine types of low invasive potential 8 years after vaccine introduction in Stockholm, Sweden. Vaccine. 2016;34(38):4565–71.

197. Jefferies JM, Johnston CH, Kirkham LA, Cowan GJ, Ross KS, Smith A, et al. Presence of nonhemolytic pneumolysin in serotypes of Streptococcus pneumoniae associated with disease outbreaks. J Infect Dis. 2007;196(6):936–44.

198. Thomas HL, Gajraj R, Slack MPE, Sheppard C, Hawkey P, Gossain S, et al. An explosive outbreak of Streptococcus pneumoniae serotype-8 infection in a highly vaccinated residential care home, England, summer 2012. Epidemiol Infect. 2015;143(9):1957–63.

199. Balsells E, Dagan R, Yildirim I, Gounder PP, Steens A, Muñoz-Almagro C, et al. The relative invasive disease potential of Streptococcus pneumoniae among children after PCV introduction: A systematic review and meta-analysis. Journal of Infection. 2018;77(5):368–78.

200. Müller A, Lekhuleni C, Hupp S, et al. Meningitis-associated pneumococcal serotype 8, ST 53, strain is hypervirulent in a rat model and has non-haemolytic pneumolysin which can be attenuated by liposomes. Front Cell Infect Microbiol. 2023.

201. Levy C, Varon E, Ouldali N, Béchet S, Bonacorsi S, Cohen R. Changes in invasive pneumococcal disease spectrum after 13-valent pneumococcal conjugate vaccine implementation. Clinical Infectious Diseases. 2020;70(3):446–54.

202. Zintgraff, J., Fossati, S., Pereira, C. S., Veliz, O., Regueira, M., Moscoloni, M. A., Irazu, M., Lara, C., Napoli D. Distribution of PCV13 and PPSV23 Streptococcus pneumoniae serotypes in Argentinean adults with invasive disease, 2013-2017. Rev Argent Microbiol. 2020;52(3):189–94.

203. Horácio AN, Silva-Costa C, Lopes JP, Ramirez M, Melo-Cristino J, Portuguese Group for the Study of Streptococcal Infections. Serotype 3 remains the leading cause of invasive pneumococcal disease in adults in Portugal (2012–2014) despite continued reductions in other 13-valent conjugate vaccine serotypes. Frontiers in Microbiology. 2016;7:1616.

204. Ardanuy, C., Adela, G., García, E., Fenoll, A., Calatayud, L., Cercenado, E., Pérez-Trallero, E., Bouza, E & Liñares J. Spread of Streptococcus pneumoniae serotype 8-ST63 multidrug-resistant recombinant Clone, Spain. Emerg Infect Dis. 2014;20(11):1848–56.

205. GERMS SA. Annual surveillance review. 2023.

206. GERMS SA. Centre for Respiratory Diseases and Meningitis Quarterly report on laboratory-based surveillance for invasive bacterial disease caused by vaccine-preventable or epidemic-prone pathogens. 2025.

Appendices

Appendix A

Ethics clearance certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Ms. Cebile Lekhuleni

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M210749

NAME: (Principal Investigator)	Ms. Cebile Lekhuleni
DEPARTMENT:	School of Pathology National Institute for Communicable Diseases Centre for Respiratory Diseases and Meningitis
PROJECT TITLE:	Molecular Characterisation and Population Structure Determination of Streptococcus pneumoniae in South Africa using Whole-Genome Analysis
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M1809107
SUPERVISOR:	Dr M. du Plessis, Dr K. Ndlangisa and Dr A. von Gottberg
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	03/08/2021

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions.Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. I **agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **July** and will therefore be due in the month of **July** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature — Date: 09 Aug 2021

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B

A look-up table for GPSC lineages, sequence types and clonal complexes . Only GPSCs with ≥10 isolates are shown (N=4314 of 4669).

GPSC	Number of isolates	Sequence types	Clonal complexes and singletons
1	124	271, 320, 420, 1412, 2706, 2834, 2920, 8678, 9933, 10683, 10731, 10822, 10825, 11129, 11696, 12134, 12136, 12228, 12229, 12526, 12533, 12534, 12536, 12537, 12543, 12560, 12586	CC271, CC2834, CC420, ST10683, ST12228, ST12229, ST12533
2	323	217, 303, 611, 612, 618, 8158, 8314, 9067, 11104, 11743, 12546, 12556	CC217, CC615
3	447	53, 62, 100, 445, 673, 1012, 1110, 2420, 3847, 10068, 10711, 10857, 11074, 11124, 11686, 12148, 12542, 12547, 12574, 17010	CC1012, CC2420, CC445, CC53, CC62, ST100, ST1110, ST11124, ST17010
4	21	199, 201, 411, 416, 2220, 3934	CC199, ST416
5	186	171, 172, 361, 1349, 1447, 1535, 2218, 4168, 9913, 10603, 10732, 10744, 10817, 12126, 12525, 12541, 12580, 14492, 17026, 17718, 17919	CC172, ST1349, ST14492, ST1535, ST17026, ST171, ST17718, ST17919, ST4168
6	19	156, 162, 312, 2722, 9910, 12554	CC156
7	42	36, 42, 355, 438, 507, 992, 2319, 10727, 11063, 15370	CC2319, CC439, ST15370, ST36
8	51	289, 5659, 7050	CC5659
9	89	63, 782, 1191, 2097, 2105, 2414, 2613, 5004, 7061, 10737, 10758, 11687, 12555, 12559, 12576, 17795, 17927	CC12576, CC63, ST17795, ST17927
10	196	230, 276, 700, 2013, 2707, 3135, 4253, 4677, 5033, 5034, 6227, 6353, 8374, 10696, 10745, 11100, 11106, 11748, 11749, 12498, 16092, 17009, 17615, 17715, 17791	CC230, ST11100, ST16092, ST17009, ST17615, ST17715, ST17791, ST2013, ST8374
11	62	193, 1262, 4893, 6237, 7068, 7723, 10065, 10774, 10783, 10854, 11760, 17151	CC11059, CC1262, CC193, CC4893, CC7723, ST10065, ST10783, ST11760, ST17151
12	26	180	CC180
13	153	471, 473, 1096, 2285, 2295, 4087, 5073, 6372, 6390, 7310, 7313, 7733, 7736, 8696, 9915, 9920, 9923, 10630, 10664, 10701, 10704, 10713, 10718, 10736, 10748, 10750, 10773, 10844, 10847, 11099, 11123, 11125, 12225, 12227, 12529, 12545, 12548, 12553, 12561, 12568, 12570, 16567, 16568	CC2285, CC4087, CC473, ST10701, ST10704, ST10713, ST10773, ST11123, ST12529, ST12570, ST16568, ST7313, ST7733, ST9923
14	201	242, 805, 1444, 1445, 2059, 2651, 5653, 6279, 6285, 6305, 8076, 8299, 8335, 8337, 8716, 8741, 9911, 9932, 10571, 10604, 10653, 10735, 10738, 10791, 10796, 10816, 10818, 10859, 12147, 12562	CC10791, CC2059, CC6279, ST10791, ST2651
16	64	66, 70, 72, 81, 282, 6216, 10869, 11683	CC66, CC81, ST70

17	326	2057, 2062, 4869, 4872, 4873, 6121, 6255, 6258, 6261, 6262, 6266, 6268, 7895, 7897, 7899, 7903, 8309, 8659, 9900, 10690, 10693, 10741, 10756, 10790, 10843, 10870, 11085, 11122, 11677, 12226, 12521, 12564, 14429, 17030	CC2062, ST12226, ST8309
18	68	13, 1492, 2652, 4191, 10747, 10799, 10814, 11675, 12232, 12552	CC1492, CC15, ST10799, ST10814
21	105	347, 5179, 6095, 6277, 8670, 9927, 10536, 10553, 10572, 10619, 10709, 10752, 10784, 10794, 10798, 10935, 12146, 12585, 12700, 16974, 17014, 17033, 17155	CC347, ST10619, ST10709, ST10752, ST10784, ST12146, ST12700, ST16974, ST17014, ST17033, ST17155, ST5179, ST8670, ST9927
22	59	4984, 5408, 5902, 8817, 10545, 10547, 10554, 10734, 10753, 10761, 10856, 10861, 11726, 11727, 11732, 12293, 12294, 12540, 12565, 12575, 12583, 17719	CC11726, CC5902, CC8817, ST10734, ST10753, ST11727, ST12540, ST12575, ST17719
24	32	138, 176, 6389, 10742	CC138, CC172
25	50	4209, 8687	CC8687
26	144	989, 5026, 11075, 11105, 17002, 17005, 17011, 17025	CC989
30	59	4084, 7062, 8714, 8774, 9174, 10555, 12133, 12245	CC12133, CC4084, CC7062, ST12133, ST12245, ST9174
32	42	218, 405, 1176, 3544, 10663, 12469, 17156	CC218, CC3544, ST17156
33	71	4088, 5072, 7888, 9568, 10543, 10548, 10592, 10667, 10800, 10853, 10862, 10950, 11055, 11072, 11090, 11111, 11772, 12137, 12550, 12551, 17796	CC4088, ST10543, ST10853, ST10862, ST12137, ST17796
34	32	7067, 10544, 10680, 10697, 10710, 10865, 12588	CC7067, ST10680, ST10865
35	11	97, 8821, 8825	CC273, CC8825
36	29	446, 2068, 10864, 17721	CC1635, CC2068, ST17721
37	85	2421, 2909, 4929, 6290, 6393, 7084, 7306, 8379, 8389, 8391, 8771, 9902, 10695, 10707, 10743, 10751, 10760, 10763, 10770, 10775, 10788, 10815, 11695, 12557	CC10743, CC2421, CC7084, ST10695, ST10751, ST10760, ST10770, ST10815, ST12557
38	22	310, 393	CC393
41	101	1094, 4941, 7251, 8690, 8766, 9909, 10656, 10691, 10708, 10729, 10749, 10785, 10820, 12538, 12566, 12569, 16888	CC1094, ST12569
46	49	30, 3450, 10584, 10673, 10733, 10858, 17617, 17634, 17744	CC30, CC3450, ST17617, ST17634, ST17744
48	61	1203, 7052	CC12127, CC346
49	27	392, 6189, 10564, 11708, 14438	CC392, ST10564
51	89	458, 11128, 12590	CC458, ST11128, ST12590
52	71	5647, 8713, 10595, 10702, 10849, 12581, 17013, 17015	CC10937, CC5647

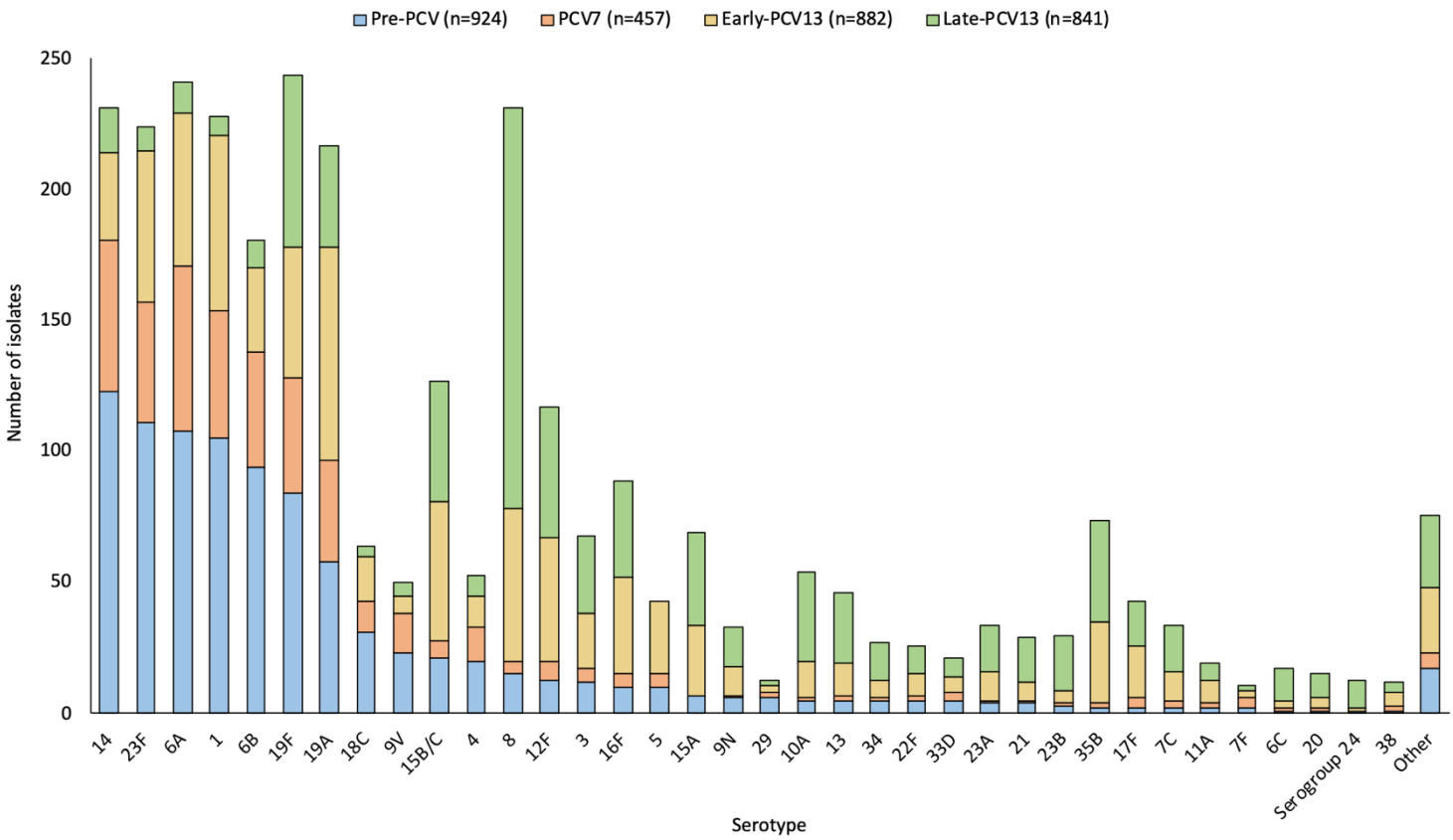
54	49	3454, 4881, 5778, 6530, 7889, 8763, 10688, 10789, 10797, 11680, 12532, 12558, 12567	CC4881, ST10797, ST12558
56	102	2416, 5027, 10792, 11767, 11768	CC2416, CC5027, ST10792
57	19	444, 1766, 3358, 3548, 10821, 12582	CC3548
61	45	6188, 6477, 8835, 9537, 10689, 10715, 11737, 11738, 11744, 12579	CC10689, CC8835, ST10715, ST6188, ST6477, ST9537
67	25	1233, 6239, 8316, 12544	CC1381, CC6239
68	39	102, 1016, 6236, 6245, 7262, 8317, 9484, 10746, 11698, 12584	CC1016, CC102, CC6236, ST11698
70	83	1221	ST1221
77	48	185, 7307, 7738, 10661, 10686, 10720, 10726, 10845, 10846, 10871, 11719, 12528	CC185, ST10726, ST10845, ST10846, ST10871, ST11719
79	31	5410, 7370, 7376, 12539, 12571	CC5410
90	23	8326, 8328, 8332, 9928, 10703, 10714, 10827, 11681, 12145, 12522	CC8328, CC8332, CC8660, ST10703, ST10714, ST10827, ST11681, ST12522
92	25	7053, 7105, 10645, 11127, 11742, 17018	CC7105
93	27	4956, 8838, 10639, 11126, 11713	CC4956, ST10639
98	10	1480, 3406, 10852	CC1480, CC3406, ST10852
102	25	4423, 7687, 17031, 17558	CC4423, CC7687, ST17031, ST17558
114	22	5326, 8827, 10787, 12530, 17773, 17778	CC5326, ST17773, ST17778
116	21	5080, 11102, 12577	CC5080
125	16	5624, 10642, 12233	CC5624, ST10642, ST12233
137	38	3983	CC12610
142	16	2419, 8840, 9523, 10588, 12549, 17768	CC10588, CC8840, ST12549, ST17768
145	10	5483, 10826	CC5483
155	13	105, 5604, 12587	CC105
160	11	9813	ST9813
168	29	10698, 11766, 17780	ST10698, ST11766, ST17780
178	33	2185, 4879, 7345, 15354	CC2185, ST15354, ST4879, ST7345
205	17	2067, 10716, 10823, 12535, 17024	CC2067

Appendix C

Case characteristics of invasive pneumococcal disease isolates stratified by age group: <5 and 5-17 years, South Africa, 2005-2020, N=17 121 (total number of reported cases), n=11 833 (total number of viable isolates), nseq=3104 (number of sequenced isolates), and % seq is the percentage of isolates sequenced.

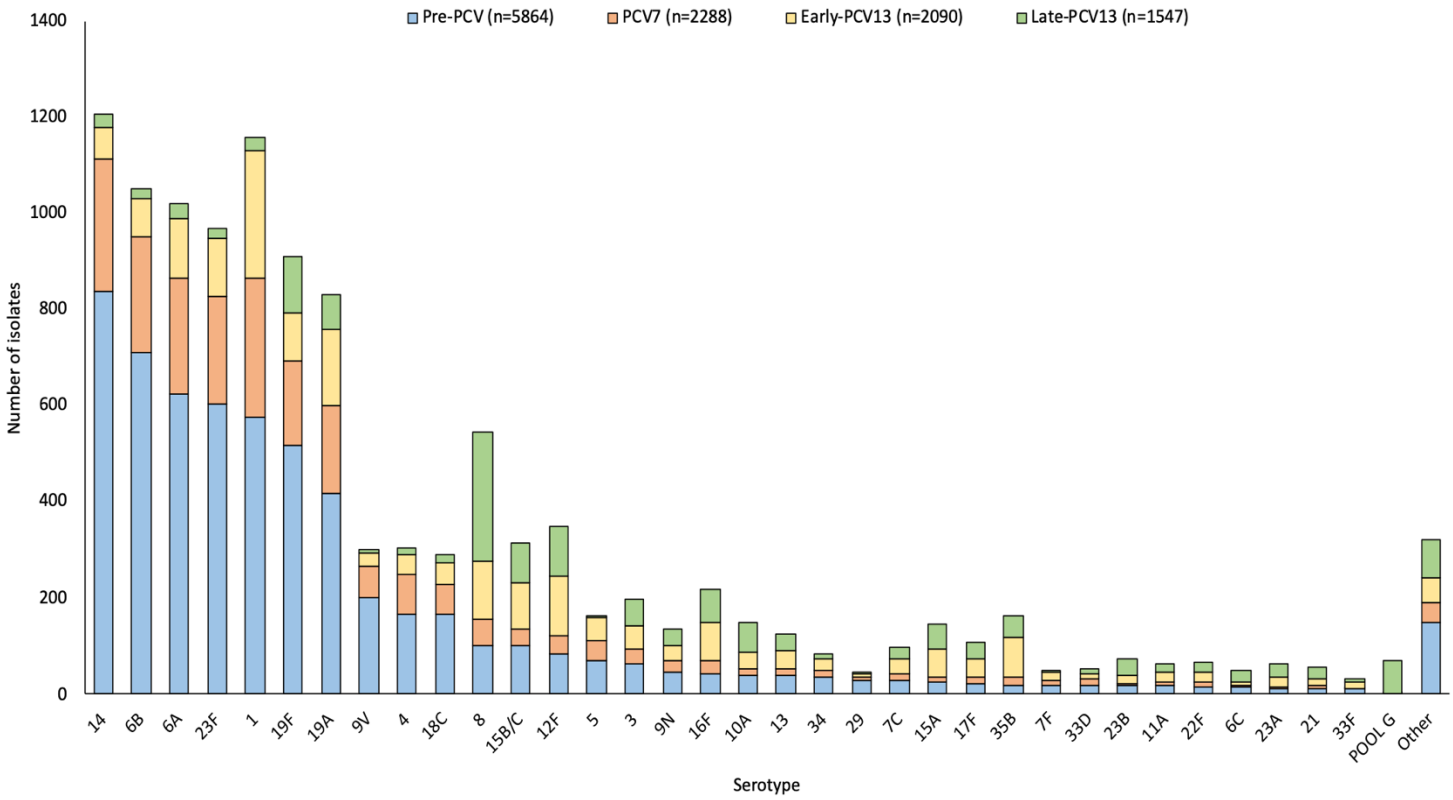
		<5 (N=12473; n=8725; nseq=2680)					5-17 (N=4648; n=3108; nseq=424)					Total (N=17121; n=11833; nseq=3104)				
		N	n viable	n/N (%)	n seq	% seq	N	n viable	n/N (%)	n seq	% seq	N	n viable	n/N (%)	n seq	% seq
Year	2005	1554	1216	78.2	187	15.4	490	382	78	32	8.4	2044	1598	78.2	219	13.7
	2006	1433	1081	75.4	197	18.2	494	371	75.1	27	7.3	1927	1452	75.4	224	15.4
	2007	1467	1084	73.9	191	17.6	436	332	76.1	53	16.0	1903	1416	74.4	244	17.2
	2008	1465	1099	75.0	197	17.9	433	313	72.3	40	12.8	1898	1412	74.4	237	16.8
	2009	1336	1009	75.5	195	19.3	496	361	72.8	31	8.6	1832	1370	74.8	226	16.5
	2010	910	650	71.4	198	30.5	393	275	70	33	12.0	1303	925	71	231	25.0
	2011	695	464	66.8	199	42.9	350	216	61.7	31	14.4	1045	680	65.1	230	33.8
	2012	512	356	69.5	206	57.9	312	201	64.4	32	15.9	824	557	67.6	238	42.7
	2013	498	322	64.7	183	56.8	225	122	54.2	23	18.9	723	444	61.4	206	46.4
	2014	465	301	64.7	188	62.5	194	119	61.3	20	16.8	659	420	63.7	208	49.5
	2015	382	216	56.5	169	78.2	162	88	54.3	15	17.0	544	304	55.9	184	60.5
	2016	400	231	57.8	169	73.2	176	90	51.1	21	23.3	576	321	55.7	190	59.2
	2017	387	210	54.3	129	61.4	136	57	41.9	11	19.3	523	267	51.1	140	52.4
	2018	386	175	45.3	34	19.4	143	60	42	18	30.0	529	235	44.4	52	22.1
	2019	361	180	49.9	143	79.4	139	81	58.3	18	22.2	500	261	52.2	161	61.7
	2020	222	131	59.0	95	72.5	69	40	58	19	47.5	291	171	58.8	114	66.7
Cases	Pre-PCV	5919	4480	75.7	772	17.2	1853	1398	75.4	152	10.9	7772	5878	75.6	924	15.7
	PCV7	2246	1659	73.9	393	23.7	889	636	71.5	64	10.1	3135	2295	73.2	457	19.9
	Early-PCV13	2170	1443	66.5	776	53.8	1081	658	60.9	106	16.1	3251	2101	64.6	882	42.0
	Late-PCV13	2138	1143	53.5	739	64.7	825	416	50.4	102	24.5	2963	1559	52.6	841	53.9
Gender	Female	5546	3850	69.4	1169	30.4	2175	1446	66.5	197	13.6	7721	5296	68.6	1366	25.8
	Male	6635	4692	70.7	1468	31.3	2387	1617	67.7	221	13.7	9022	6309	69.9	1689	26.8
	Unknown	292	183	62.7	43	23.5	86	45	52.3	6	13.3	378	228	60.3	49	21.5
Serotype	PCV7	4029	3955	98.2	908	23.0	1129	1098	97.3	139	12.7	5158	5053	98	1047	20.7
	PCV13 unique serotypes	2306	2217	96.1	647	29.2	1272	1204	94.7	161	13.4	3578	3421	95.6	808	23.6
	Non-PCV serotypes	2930	2553	87.1	1125	44.1	916	806	88	124	15.4	3846	3359	87.3	1249	37.2
	Non-viable	3208	-	-	-	-	1331	-	-	-	-	4539	-	-	-	-
Specimen	Cerebrospinal fluid	4080	2707	66.3	777	28.7	2057	1222	59.4	164	13.4	6137	3929	64	941	24.0
	Blood	7927	5763	72.7	1807	31.4	2347	1778	75.8	249	14.0	10274	7541	73.4	2056	27.3
	Other	466	255	54.7	96	37.6	244	108	44.3	11	10.2	710	363	51.1	107	29.5
Diagnosis	Meningitis	2651	2111	79.6	677	32.1	1124	871	77.5	112	12.9	3775	2982	79	789	26.5
	Bacteraemia without focus	530	425	80.2	136	32.0	121	94	77.7	13	13.8	651	519	79.7	149	28.7
	Lower respiratory tract infection	2763	2260	81.8	670	29.6	858	700	81.6	99	14.1	3621	2960	81.7	769	26.0
	Other	447	336	75.2	102	30.4	113	75	66.4	9	12.0	560	411	73.4	111	27.0
	Unknown	6082	3593	59.1	1095	30.5	2432	1368	56.3	191	14.0	8514	4961	58.3	1286	25.9

Appendix D



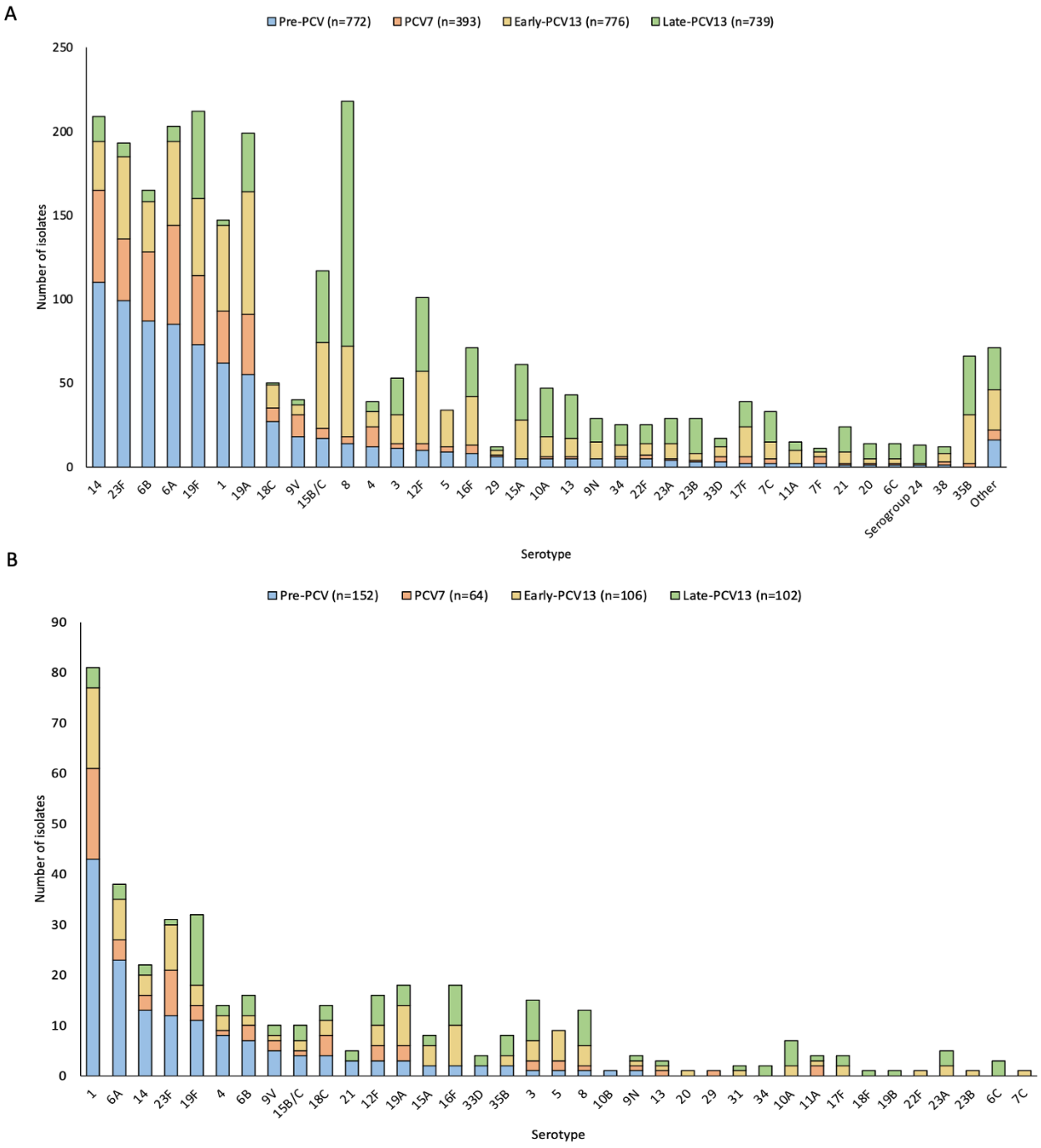
Serotype distribution among invasive pneumococcal disease isolates in individuals <18 years in South Africa (N=3104) across the pre-PCV (2005-2008), PCV7 (2009-2010), early-PCV13 (2011-2014), and late-PCV13 (2015-2020) periods. Serotypes are listed in order of pre-PCV prevalence. Other=serotypes with <10 isolates.

Appendix E



Serotype distribution among all serotypes (N=11 789) causing invasive pneumococcal disease in individuals <18 years in South Africa across the pre-PCV (2005-2008), PCV7 (2009-2010), early-PCV13 (2011-2014), and late-PCV13 (2015-2020) periods. Serotypes are listed in order of overall prevalence. Other=serotypes with <30 isolates. Only isolates with identifiable serotypes were included; dual serotype infection isolates were excluded.

Appendix F



Serotype distribution among invasive pneumococcal disease isolates from children <5 years (A, n=2680) and individuals aged 5-17year (B, n=424) in South Africa across the pre-PCV (2005-2008), PCV7 (2009-2010), early-PCV13 (2011-2014), and late-PCV13 (2015-2020) periods. Serotypes are listed in order of pre-PCV prevalence. Other=serotypes with <10 isolates.

Appendix G

Incidence rate ratios (IRRs) of GPSC lineages comparing the pre-PCV and PCV7 periods among South African children (<5 years), 2005-2020.

GPSC	Model	pre-avg-incidence	post-avg-incidence	IRR_average	lower	upper	adj.avg.p
1	negative bionomial	1.143	0.831	0.722	0.490	1.063	0.238
2	negative bionomial	2.355	1.614	0.683	0.519	0.897	0.036
3	negative bionomial	0.725	0.332	0.443	0.252	0.778	0.026
5	negative bionomial	1.124	1.237	1.086	0.764	1.544	0.729
7	none	0.188	0.045	0.182	0.040	0.831	0.055
8	negative bionomial	0.348	0.125	0.354	0.148	0.847	0.055
9	negative bionomial	0.782	0.440	0.546	0.329	0.906	0.067
10	negative bionomial	2.070	1.297	0.630	0.467	0.851	0.026
11	none	0.410	0.126	0.303	0.129	0.714	0.026
13	negative bionomial	1.656	1.135	0.683	0.492	0.946	0.067
14	negative bionomial	2.254	1.488	0.655	0.493	0.869	0.026
17	negative bionomial	1.938	1.647	0.854	0.643	1.136	0.424
18	negative bionomial	1.165	0.909	0.771	0.529	1.124	0.328
21	poisson robust SE	0.914	0.738	0.811	0.532	1.236	0.444
22	poisson robust SE	0.148	0.081	0.455	0.137	1.511	0.356
24	negative bionomial	0.491	0.433	0.874	0.499	1.530	0.729
30	none	0.109	0.162	1.365	0.486	3.835	0.709
32	negative bionomial	0.068	0.172	2.275	0.714	7.254	0.328
33	none	0.109	0.162	1.365	0.486	3.835	0.709
34	none	0.158	0.063	0.455	0.137	1.511	0.356
37	negative bionomial	1.258	1.091	0.853	0.600	1.213	0.516
41	negative bionomial	1.458	1.486	1.008	0.737	1.381	1.000
48	negative bionomial	0.412	0.108	0.260	0.105	0.644	0.026
51	negative bionomial	0.312	0.127	0.398	0.164	0.968	0.094
52	poisson robust SE	0.346	0.108	0.303	0.120	0.764	0.036
54	negative bionomial	0.653	0.469	0.717	0.429	1.199	0.356
56	negative bionomial	0.256	0.271	1.050	0.500	2.207	1.000
67	negative bionomial	0.342	0.235	0.696	0.338	1.433	0.464
68	negative bionomial	0.406	0.172	0.455	0.213	0.972	0.101
70	none	0.109	0.369	3.034	1.218	7.554	0.055
77	negative bionomial	1.009	0.171	0.178	0.091	0.351	<0.001
79	negative bionomial	0.258	0.252	0.980	0.461	2.085	1.000
90	negative bionomial	0.383	0.125	0.335	0.141	0.798	0.042
113	negative bionomial	0.160	0.064	0.455	0.137	1.511	0.356
142	none	0.148	0.063	0.455	0.137	1.511	0.356

145	negative bionomial	0.147	0.045	0.228	0.048	1.071	0.121
205	negative bionomial	0.243	0.108	0.455	0.171	1.212	0.302

Appendix H

IRRs of GPSC lineages comparing the pre-PCV and PCV7 periods among South African children (5-17 years), 2005-2020.

GPSC	model	pre-avg-incidence	post-avg-incidence	IRR_average	lower	upper	adj.avg.p
1	none	0.115	0.096	0.821	0.384	1.753	0.702
2	negative bionomial	1.019	0.954	0.940	0.734	1.204	0.702
9	poisson	0.146	0.008	0.054	0.007	0.403	<0.001
10	none	0.154	0.128	0.821	0.425	1.584	0.702
13	negative bionomial	0.366	0.064	0.175	0.083	0.370	<0.001
14	negative bionomial	0.165	0.575	3.517	2.163	5.719	<0.001
17	negative bionomial	0.072	0.112	1.596	0.691	3.687	0.665
21	none	0.095	0.064	0.684	0.280	1.673	0.702
41	none	0.148	0.112	0.756	0.379	1.508	0.702
54	none	0.103	0.048	0.473	0.180	1.246	0.461
68	negative bionomial	0.094	0.112	1.197	0.554	2.588	0.702

Appendix I

IRRs of GPSC lineages comparing the pre-PCV and late-PCV13 periods among South African children (<5 years), 2005-2020.

GPSC	model	pre-avg-incidence	post-avg-incidence	IRR_average	lower	upper	adj.avg.p
1	negative bionomial	1.144	0.203	0.184	0.099	0.343	<0.001
2	none	2.355	0.021	0.007	0.001	0.053	<0.001
3	negative bionomial	0.725	1.317	1.805	1.218	2.677	0.008
4	negative bionomial	0.077	0.029	0.445	0.082	2.431	0.535
5	negative bionomial	1.124	0.417	0.375	0.233	0.604	<0.001
7	negative bionomial	0.190	0.059	0.267	0.074	0.971	0.087
8	poisson robust SE	0.366	0.018	0.047	0.006	0.350	<0.001
9	negative bionomial	0.781	0.138	0.178	0.083	0.381	<0.001
10	negative bionomial	2.072	0.291	0.137	0.081	0.232	<0.001
11	none	0.410	0.161	0.382	0.175	0.833	0.032
12	negative bionomial	0.041	0.038	0.891	0.125	6.323	1.000
13	negative bionomial	1.656	0.044	0.021	0.005	0.086	<0.001
14	negative bionomial	2.254	0.067	0.031	0.012	0.085	<0.001
16	negative bionomial	0.130	0.094	0.742	0.227	2.432	0.893
17	negative bionomial	1.939	0.270	0.136	0.079	0.235	<0.001
18	poisson	1.186	0.018	0.015	0.002	0.107	<0.001
21	negative bionomial	0.919	0.200	0.213	0.110	0.411	<0.001
22	poisson	0.148	0.132	0.891	0.334	2.373	1.000
24	poisson	0.509	0.018	0.034	0.005	0.252	<0.001
25	poisson robust SE	0.020	0.076	3.563	0.398	31.875	0.506
26	negative bionomial	0.151	0.246	1.559	0.654	3.715	0.514
30	poisson robust SE	0.109	0.126	1.039	0.349	3.092	1.000
33	none	0.109	0.056	0.445	0.111	1.781	0.440
34	negative bionomial	0.165	0.050	0.334	0.089	1.259	0.207
36	none	0.079	0.094	1.113	0.299	4.146	1.000
37	negative bionomial	1.257	0.053	0.042	0.013	0.133	<0.001
38	negative bionomial	0.033	0.029	0.891	0.125	6.323	1.000
41	none	1.477	0.018	0.012	0.002	0.085	<0.001
46	negative bionomial	0.074	0.173	2.227	0.698	7.100	0.287
48	negative bionomial	0.412	0.155	0.382	0.175	0.833	0.032
49	negative bionomial	0.074	0.047	0.668	0.150	2.985	0.850
51	negative bionomial	0.312	0.091	0.278	0.102	0.760	0.018
52	negative bionomial	0.347	0.144	0.396	0.172	0.910	0.057
54	negative bionomial	0.653	0.021	0.027	0.004	0.197	<0.001
56	negative bionomial	0.256	0.088	0.343	0.122	0.961	0.066

57	none	0.074	0.032	0.445	0.082	2.431	0.535
61	poisson	0.020	0.111	5.344	0.643	44.389	0.207
67	poisson	0.361	0.018	0.049	0.007	0.371	<0.001
68	negative bionomial	0.406	0.006	0.042	0.006	0.315	<0.001
70	negative bionomial	0.106	0.018	0.148	0.018	1.233	0.097
77	negative bionomial	1.009	0.012	0.017	0.002	0.126	<0.001
79	negative bionomial	0.258	0.006	0.064	0.008	0.484	<0.001
90	negative bionomial	0.383	0.029	0.094	0.022	0.403	<0.001
92	negative bionomial	0.041	0.103	2.672	0.539	13.239	0.414
98	negative bionomial	0.035	0.053	1.336	0.223	7.996	1.000
102	negative bionomial	0.020	0.226	11.579	1.515	88.511	0.006
114	negative bionomial	0.107	0.038	0.297	0.060	1.471	0.248
116	negative bionomial	0.039	0.067	1.781	0.326	9.726	0.841
125	none	0.124	0.018	0.148	0.018	1.233	0.097
126	none	0.040	0.056	1.336	0.223	7.996	1.000
137	poisson	0.020	0.088	4.453	0.520	38.119	0.321
142	negative bionomial	0.146	0.006	0.099	0.013	0.781	0.018
145	negative bionomial	0.147	0.006	0.099	0.013	0.781	0.018
168	poisson	0.020	0.091	4.453	0.520	38.119	0.321
178	none	0.040	0.041	0.891	0.125	6.323	1.000
205	negative bionomial	0.243	0.018	0.074	0.010	0.571	0.003

Appendix J

IRRs of GPSC lineages comparing the pre-PCV and late-PCV13 periods among South African children (5-17 years), 2005-2020.

GPSC	model	pre-avg-incidence	post-avg-incidence	IRR_average	lower	upper	adj.avg.p
1	negative bionomial	0.115	0.045	0.375	0.146	0.967	0.085
2	negative bionomial	1.019	0.029	0.029	0.011	0.077	<0.001
3	none	0.037	0.055	1.501	0.491	4.590	0.808
5	none	0.101	0.024	0.217	0.062	0.760	0.025
9	negative bionomial	0.138	0.010	0.052	0.007	0.391	<0.001
10	negative bionomial	0.153	0.050	0.328	0.139	0.777	0.025
11	none	0.053	0.058	1.072	0.389	2.957	1.000
13	negative bionomial	0.366	0.014	0.040	0.010	0.164	<0.001
14	negative bionomial	0.165	0.024	0.134	0.040	0.449	<0.001
17	none	0.072	0.044	0.626	0.223	1.758	0.665
21	none	0.095	0.077	0.782	0.338	1.810	0.814
26	none	0.029	0.041	1.408	0.397	4.988	0.814
30	none	0.045	0.038	0.782	0.239	2.562	0.814
33	none	0.035	0.047	1.408	0.397	4.988	0.814
37	none	0.086	0.023	0.256	0.071	0.917	0.061
41	none	0.156	0.007	0.047	0.006	0.350	<0.001
51	none	0.021	0.049	2.190	0.566	8.467	0.567
79	negative bionomial	0.071	0.010	0.104	0.013	0.823	0.025

Appendix K

Antibiotic resistance mechanisms within study pneumococcal isolates (N=3104) among South African children <18 years, 2005-20. NS=nonsusceptible. S=susceptible. Neg=negative for resistance determinant. Penicillin nonsusceptibility was predicted based on the penicillin-binding protein (PBP) types: *pbp*1a, *pbp*2x, *pbp*2b; chloramphenicol, based on the presence of chloramphenicol acetyltransferase gene, *cat*; erythromycin, based on the presence of erythromycin resistance methylase gene *ermB* or macrolide efflux pump gene *mefA*; tetracycline, by the presence of *tet*M or *tet*S/M gene with no promoter region interruptions; cotrimoxazole, by the presence of mutation I100L in *folA* and/or indel within amino acid residue 56-67 in *folP.* Concordance between phenotypic and genotypic resistance phenotypes was determined. † PBP type combinations conferring penicillin nonsusceptibility.

Antibiotic	Phenotypic susceptibility	Number of isolates (N=3104)	Resistance determinants, n (%)					Overall concordance (%)
			Beta lactam resistance mutations					
Penicillin			*pbp*:1a-2b-2x†	NF	neg			95%
	NS	1350	1295 (96%)	2 (0.1%)	53 (4%)			
	S	1754	116 (7%)	0 (0%)	1638 (93%)			
			Macrolide resistance genes					
Erythromycin			*ermB*	*mefA*	*ermB*/*mef*A	neg		99%
	NS	539	288 (53%)	144 (27%)	77 (14%)	30 (6%)		
	S	2565	6 (0·2%)	2 (0·01%)	2 (0·1%)	2555 (100%)		
			Lincosamide resistance gene					
Clindamycin			*ermB*	*ermB*/*mef*A	neg			99%
	NS	371	287	70 (19%)	14 (4%)			
	S	2733	7	9 (0·3%)	2717 (99%)			
			Chloramphenicol resistance gene					
Chloramphenicol			*cat*_pC194	neg				100%
	NS	128	119 (93%)	9 (7%)				
	S	2976	5 (0·2%)	2971 (100%)				
			Tetracyclines resistance genes					
Tetracycline			*tet*M	*tet*S_M	neg			98%
	NS	638	525 (82%)	87 (14%)	26 (4%)			
	S	2466	47 (2%)	2 (0·1%)	2417 (98%)			
			Sulfonamides resistance mutations					
Cotrimoxazole			*fol*A_I100L	*fol*P_aa_insert_57-70	*fol*A_I100L, *fol*P_aa_insert_57-70	Flag	neg	93%
	NS	1895	28 (1%)	413 (22%)	1386 (73%)	23 (1%)	45 (2)	
	S	1209	6 (0·5%)	144 (12%)	14 (1%)	5 (0·4%)	1040 (86)	

Appendix L

Characteristics of invasive pneumococcal disease cases, South Africa, 2005–2020. N=54,199 (total number of reported cases including unknown age), n=36,546 (total number of viable isolates including unknown age), n_seq=4669 (total number of sequenced isolates with known age).

	<5 (n_viable=8722; n_seq=2680)			≥5 (n_viable=26694; n_seq=1989)			Unknown age (n_viable=1130; n_seq=na)			Total (n_viable=36546; n_seq=4669)		
	n_viable	n_seq	% seq (n_seq/n_viable)	n_viable	n_seq	% seq (n_seq/n_viable)	n_viable	n_seq	% seq (n_seq/n_viable)	n_viable	n_seq	% seq (n_seq/n_viable)
Year												
2005	1215	187	15.4	2297	98	4.3	137	na	na	3649	285	7.8
2006	1081	197	18.2	2197	97	4.4	141	na	na	3419	294	8.6
2007	1084	191	17.6	2106	131	6.2	135	na	na	3325	322	9.7
2008	1098	197	17.9	2079	96	4.6	148	na	na	3325	293	8.8
2009	1009	195	19.3	2271	92	4.1	104	na	na	3384	287	8.5
2010	650	198	30.5	2133	100	4.7	89	na	na	2872	298	10.4
2011	464	199	42.9	1880	95	5.1	65	na	na	2409	294	12.2
2012	356	206	57.9	1721	124	7.2	84	na	na	2161	330	15.3
2013	322	183	56.8	1551	79	5.1	60	na	na	1933	262	13.6
2014	301	188	62.5	1401	95	6.8	48	na	na	1750	283	16.2
2015	216	169	78.2	1437	127	8.8	48	na	na	1701	296	17.4
2016	231	169	73.2	1331	127	9.5	15	na	na	1577	296	18.8
2017	210	129	61.4	1307	154	11.8	16	na	na	1533	283	18.5
2018	175	34	19.4	1148	246	21.4	12	na	na	1335	280	21.0
2019	180	143	79.4	1193	138	11.6	12	na	na	1385	281	20.3
2020	130	95	73.1	642	190	29.6	16	na	na	788	285	36.2

Vaccine period												
Pre-PCV	4478	772	17.2	8679	422	4.9	561	na	na	13718	1194	8.7
PCV7	1659	393	23.7	4404	192	4.4	193	na	na	6256	585	9.4
Early-PCV13	1443	776	53.8	6553	393	6.0	257	na	na	8253	1169	14.2
Late-PCV13	1142	739	64.7	7058	982	13.9	119	na	na	8319	1721	20.7
Sex												
Female	3849	1169	30.4	13865	1000	7.2	485	na	na	18199	2169	11.9
Male	4690	1468	31.3	12549	972	7.7	490	na	na	17729	2440	13.8
Unknown	183	43	23.5	280	17	6.1	155	na	na	618	60	9.7
Specimen type												
Blood	5761	1807	31.4	15541	1210	7.8	443	na	na	21745	3017	13.9
Cerebrospinal fluid	2706	777	28.7	9164	660	7.2	564	na	na	12434	1437	11.6
Other	255	96	37.6	1989	119	6.0	123	na	na	2367	215	9.1
Serotype												
PCV7	3956	908	23.0	6532	417	6.4	374	na	na	10862	1325	12.2
PCV13 unique serotypes	2217	647	29.2	8680	572	6.6	355	na	na	11252	1219	10.8
Non-PCV serotypes	2549	1125	44.1	11482	1000	8.7	401	na	na	14432	2125	14.7
No isolate available	n/a	n/a	n/a	n/a	n/a	n/a	na	na	na	na	n/a	na
Nonsusceptibility profile												
Fully susceptible	2523	861	34.1	13155	980	7.4	487	na	na	16165	1841	11.4
Penicillin NS only	189	23	12.2	267	21	7.9	27	na	na	483	44	9.1
Cotrimoxazole NS only	1320	442	33.5	3694	290	7.9	158	na	na	5172	732	14.2
Penicillin-cotrimoxazole NS	2245	749	33.4	4310	338	7.8	223	na	na	6778	1087	16.0
Penicillin MDR	1919	467	24.3	3125	212	6.8	129	na	na	5173	679	13.1
Other MDR	243	105	43.2	955	112	11.7	49	na	na	1247	217	17.4
Other resistance	283	33	11.7	1188	36	3.0	57.0	na	na	1528	69	4.5

n_viable - number of viable isolates with susceptibility data

n_seq - number of sequenced isolates

% seq - percentage of sequenced isolates (number of sequenced isolates divided by number of viable isolates with susceptibility data)

n_ns - number of nonsusceptible isolates

n_seq_ns - number of sequenced nonsusceptible isolates

% seq_ns - percentage of sequenced nonsusceptible isolates (number of sequenced nonsusceptible isolates divided by number of nonsusceptible isolates)

NS = nonsusceptible

Penicillin multidrug resistance (MDR) was defined as nonsusceptibility to penicillin and two or more other antibiotic classes (erythromycin, chloramphenicol, tetracycline, rifampicin and cotrimoxazole)

Other MDR was defined as nonsusceptibility to three or more antibiotic classes.

For sequenced isolates, MDR does not include rifampicin as resistant determinant is undefined

No isolate available - include audit cases, culture negative specimens serotyped using PCR, missing and non-viable isolates.

Appendix M

Antibiotic-nonsusceptible IPD by age group (<5 and ≥5 years) and vaccine period (pre-PCV [2005–2008], PCV7 [2009–2011], early-PCV13 [2011–2014], and late-PCV13 [2015–2020] periods), South Africa, 2005-2020.

Resistance profile	<5y (N=8722)				≥5y (N=26694)			
	Pre-PCV (n=4478)	PCV7 (n=1659)	Early-PCV13 (n=1443)	Late-PCV13 (n=1142)	Pre-PCV (n=8679)	PCV7 (n=4404)	Early-PCV13 (n=6553)	Late-PCV13 (n=7058)
FULLY_SUSCEPTIBLE	1055 (%)	420 (%)	550 (%)	498 (%)	4306 (%)	2088 (%)	3205 (%)	3556 (%)
NVT	418 (39.6)	162 (38.6)	364 (66.2)	447 (89.8)	1826 (42.4)	847 (40.6)	1720 (53.7)	2693 (75.7)
VT	637 (60.4)	258 (61.4)	186 (33.8)	51 (10.2)	2480 (57.6)	1241 (59.4)	1485 (46.3)	863 (24.3)
PEN	122 (%)	27 (%)	27 (%)	13 (%)	115 (%)	32 (%)	56 (%)	64 (%)
NVT	3 (2.5)	5 (18.5)	14 (51.9)	12 (92.3)	7 (6.1)	5 (15.6)	18 (32.1)	44 (68.8)
VT	119 (97.5)	22 (81.5)	13 (48.1)	1 (7.7)	108 (93.9)	27 (84.4)	38 (67.9)	20 (31.3)
COT	751 (%)	156 (%)	238 (%)	175 (%)	1303 (%)	445 (%)	944 (%)	1002 (%)
NVT	132 (17.6)	47 (30.1)	170 (71.4)	157 (89.7)	372 (28.5)	199 (44.7)	571 (60.5)	811 (80.9)
VT	619 (82.4)	109 (69.9)	68 (28.6)	18 (10.3)	931 (71.5)	246 (55.3)	373 (39.5)	191 (19.1)
PEN_COT	1135 (%)	526 (%)	324 (%)	260 (%)	1230 (%)	877 (%)	1088 (%)	1115 (%)
NVT	52 (4.6)	26 (4.9)	89 (27.5)	148 (56.9)	63 (5.1)	46 (5.2)	172 (15.8)	321 (28.8)
VT	1083 (95.4)	500 (95.1)	235 (72.5)	112 (43.1)	1167 (94.9)	831 (94.8)	916 (84.2)	794 (71.2)
PEN MDR	1110 (%)	481 (%)	198 (%)	130 (%)	1009 (%)	651 (%)	734 (%)	731 (%)
NVT	10 (0.9)	9 (1.9)	25 (12.6)	62 (47.7)	20 (2)	16 (2.5)	93 (12.7)	294 (40.2)
VT	1100 (99.1)	472 (98.1)	173 (87.4)	68 (52.3)	989 (98)	635 (97.5)	641 (87.3)	437 (59.8)
OTHER MDR	136 (%)	30 (%)	53 (%)	24 (%)	281 (%)	130 (%)	271 (%)	273 (%)
NVT	9 (6.6)	8 (26.7)	35 (66)	22 (91.7)	78 (27.8)	55 (42.3)	209 (77.1)	258 (94.5)
VT	127 (93.4)	22 (73.3)	18 (34)	2 (8.3)	203 (72.2)	75 (57.7)	62 (22.9)	15 (5.5)

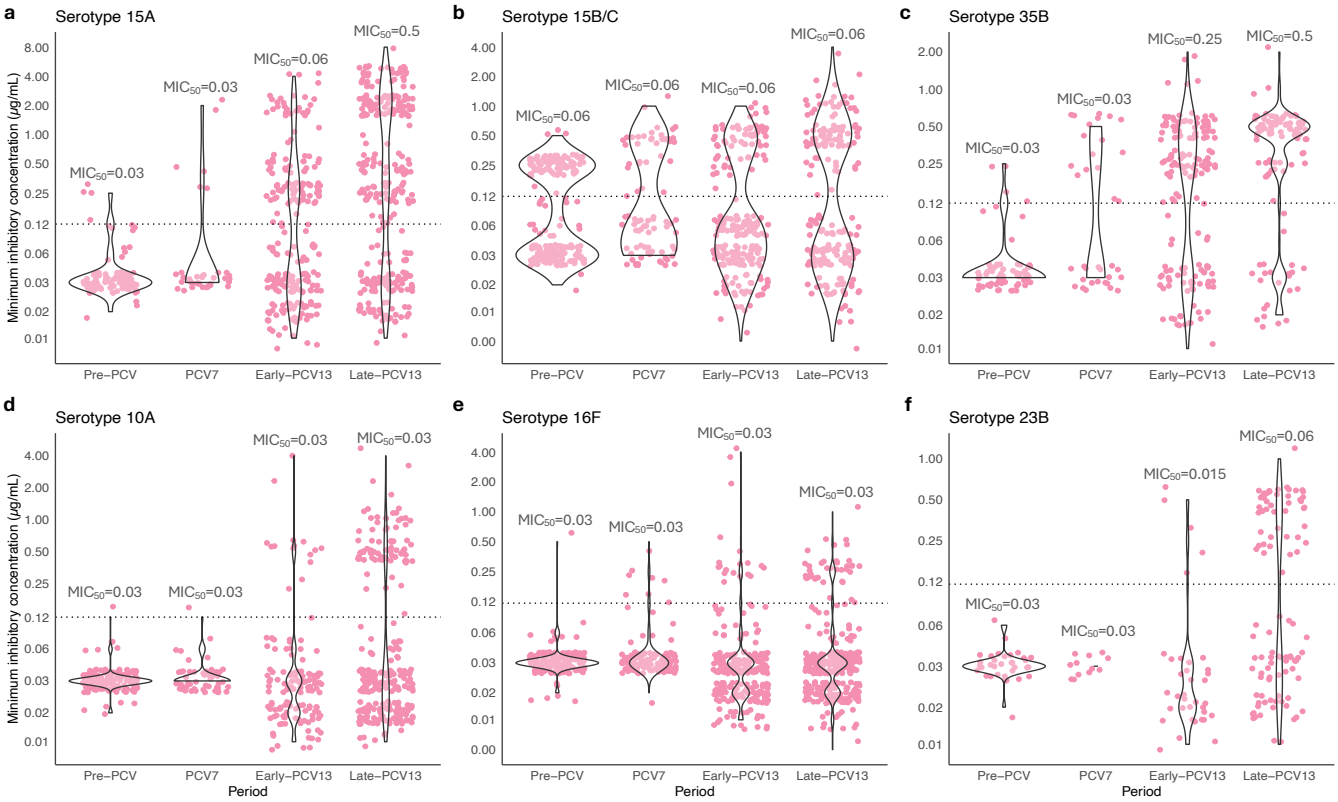
OTHER RESISTANCE	169 (%)	19 (%)	53 (%)	42 (%)	435 (%)	181 (%)	255 (%)	317 (%)
NVT	34 (20.1)	12 (63.2)	37 (69.8)	40 (95.2)	203 (46.7)	103 (56.9)	160 (62.7)	277 (87.4)
VT	135 (79.9)	7 (36.8)	16 (30.2)	2 (4.8)	232 (53.3)	78 (43.1)	95 (37.3)	40 (12.6)

NVT - non-PCV13 serotypes

VT - PCV13 serotypes

FULLY_ SUSCEPTIBLE = susceptible to all antibiotics of interest. PEN = penicillin-nonsusceptible (NS), COT = cotrimoxazole NS, PEN_COT = penicillin-cotrimoxazole NS, PEN_MDR = nonsusceptibility to penicillin plus two or more antibiotic classes, OTHER_MDR = nonsusceptibility to three or more antibiotic classes, OTHER_RES = nonsusceptibility to at most two of the following antibiotic classes: erythromycin, chloramphenicol, tetracycline, and rifampicin.

Appendix N



Scatter plot of penicillin minimum inhibitory concentrations (MIC) by vaccine period in South Africa. **a–f** shows the distribution of MICs across the pre-PCV (2005–2008), PCV7 (2009–2010), early-PCV13 (2011–2014), and late-PCV13 (2015–2020) periods. Penicillin nonsusceptibility was defined at MIC ≥0.12 μg/mL and is represented by the dotted horizontal line (meningitis breakpoint).

Appendix O

Penicillin-resistant Global Pneumococcal Sequence Clusters (GPSC) in order of prevalence, and their expressed *in silico* serotypes during the pre-PCV period (2005–2008) in all ages, South Africa.

GPSC Serotype	Number of isolates
17	**71**
19A	68
19F	3
10	**65**
14	64
19A	1
41	**49**
23B	1
6A	48
5	**45**
19A	6
23F	36
6A	3
14	**43**
23F	38
6B	5
1	**42**
19F	42
18	**36**
14	36
37	**32**
6B	32
77	**31**
6A	1
6B	30
9	**29**
14	29
21	**29**
19A	1
19F	28
13	**17**
6A	16
6B	1
79	**14**
4	14
48	**14**
15B/C	12
19F	2

54	**12**
9V	12
90	**10**
14	1
19A	1
19F	8
Other GPSCs	**44**
Total	**583**

Appendix P

Penicillin-resistant Global Pneumococcal Sequence Clusters (GPSC) in order of prevalence, and their expressed *in silico* serotypes during the pre-PCV period (2015–2020) in all ages, South Africa.

GPSC Serotype	Number of isolates
17	**111**
3	1
19A	110
10	**72**
3	21
14	17
24	10
10A	17
19A	4
19F	1
23A	2
5	**67**
15A	1
15B/C	1
19F	7
23A	1
23B	6
35B	46
35D	3
6A	2
9	**38**
14	7
15A	30
15B/C	1
21	**36**
19F	36
1	**30**
19A	1
19F	29
48	**28**
15B/C	28
14	**20**
19A	2
23F	16
6A	1
6B	1
168	**15**
15A	15

102	**14**
19F	1
23B	13
Other GPSCs	**88**
Total	**519**

Appendix Q

Resistance phenotypes of the most common *in silico* serotypes causing nonsusceptible IPD in the late-PCV13 period. Each serotype is broken down by the global pneumococcal sequence clusters (GPSC) expressing it and respective sequence types shown by vaccine periods.

Serotype	GPSC	Sequence type	Pre-PCV (2005-2008)	PCV7 (2009-2010)	Early-PCV13 (2011-2014)	Late-PCV13 (2015-2020)	Total
		Resistance phenotype					
10A	GPSC36	2068	2	1	8	14	25
		COT	2	1	8	12	23
		PEN_COT				2	2
		10864			1		1
		COT			1		1
		17721				1	1
		COT				1	1
	GPSC10	3135				11	11
		PEN MDR				11	11
		12498			1		1
		PEN MDR			1		1
		17009				1	1
		PEN MDR				1	1
		17615				3	3
		PEN MDR				3	3
		17715				2	2
		PEN MDR				2	2
	GPSC30	12245			1	12	13
		COT			1	12	13
	GPSC35	97			2		2
		FULLY SUS			2		2
		8821	2		3	2	7
		FULLY SUS	2		3	2	7
		8825			2		2
		FULLY SUS			2		2
	GPSC425	12531	1				1
		FULLY SUS	1				1
		17722				2	2
		FULLY SUS				2	2

	GPSC194	5578			2		2
		FULLY SUS			2		2
		10866			1		1
		FULLY SUS			1		1
	GPSC355	5329				1	1
		PEN_COT				1	1
		10685	1				1
		FULLY SUS	1				1
	GPSC449	10824		1			1
		FULLY SUS		1			1
		17748				1	1
		FULLY SUS				1	1
	GPSC421	10687	1				1
		FULLY SUS	1				1
	GPSC40	7054				1	1
		PEN				1	1
15A	GPSC9	63			3	7	10
		PEN MDR			3	7	10
		2105				1	1
		PEN MDR				1	1
		2613			2	20	22
		PEN MDR			2	20	22
		12576			1		1
		PEN MDR			1		1
		17795				1	1
		PEN MDR				1	1
		17927				1	1
		PEN MDR				1	1
	GPSC168	10698	1				1
		PEN_COT	1				1
		11766			13	14	27
		PEN_COT			13	14	27
		17780				1	1
		PEN_COT				1	1
	GPSC22	10554				1	1
		COT				1	1
		10761	1				1
		COT	1				1
		10861			1		1
		COT			1		1
		11726	1		2		3
		COT			2		2

		FULLY SUS	1				1
		11727			1		1
		COT			1		1
		11732	1		1	2	4
		COT	1		1	2	4
		12293				1	1
		PEN_COT				1	1
		12294				2	2
		COT				2	2
		17719				1	1
		COT				1	1
	GPSC3	2420	1		2	4	7
		COT	1		2	4	7
		17010				1	1
		COT				1	1
	GPSC179	10550			5	2	7
		COT				1	1
		FULLY SUS			5	1	6
	GPSC142	10588	2		1		3
		FULLY SUS	2		1		3
		17768				1	1
		FULLY SUS				1	1
	GPSC133	12578			1	1	2
		COT			1	1	2
	GPSC40	10605			1	1	2
		COT			1	1	2
	GPSC5	17026				1	1
		PEN_COT				1	1
	GPSC218	12524	1				1
		FULLY SUS	1				1
	GPSC11	10065	1				1
		FULLY SUS	1				1
	GPSC140	58			1		1
		FULLY SUS			1		1
	GPSC278	7069				1	1
		PEN_COT				1	1
	GPSC69	4965	1				1
		FULLY SUS	1				1
	GPSC102	7687				1	1
		COT				1	1
15B/C	GPSC48	7052	12	2	17	28	59
		PEN_COT	12	2	17	28	59

	GPSC25	4209		1	4		5
		COT		1	3		4
		PEN_COT			1		1
		8687	1	6	26	12	45
		COT	1	6	26	12	45
	GPSC4	199			3	2	5
		FULLY SUS			3	2	5
		201	1		2	1	4
		FULLY SUS	1		2	1	4
		411	3		1	2	6
		FULLY SUS	3		1	2	6
		2220				1	1
		FULLY SUS				1	1
		3934	1				1
		FULLY SUS	1				1
	GPSC11	1262			3	6	9
		COT			3	6	9
	GPSC200	5808	3				3
		PEN_COT	3				3
		10617	1		1		2
		PEN_COT	1		1		2
	GPSC133	11763	1	1	2		4
		COT	1	1	2		4
	GPSC132	3280				2	2
		PEN MDR				2	2
		7479				1	1
		PEN MDR				1	1
	GPSC429	8820	1				1
		COT	1				1
		17726				1	1
		COT				1	1
	GPSC201	11774				1	1
		COT				1	1
	GPSC430	10739	1				1
		FULLY SUS	1				1
	GPSC9	2613				1	1
		PEN MDR				1	1
	GPSC6	312				1	1
		COT				1	1
	GPSC5	10603				1	1
		PEN_COT				1	1
	GPSC131	3669			1		1

		COT			1		1
	GPSC90	8326			1		1
		PEN_COT			1		1
16F	GPSC33	4088	6	1	5	9	21
		COT	2		1	1	4
		FULLY SUS	4	1	4	8	17
		5072			1	1	2
		COT			1	1	2
		7888	1	1	4		6
		FULLY SUS	1	1	4		6
		9568				2	2
		FULLY SUS				2	2
		10543		2	3	5	10
		PEN_COT		2	3	5	10
		10548			2		2
		COT			2		2
		10592	1		4	4	9
		COT	1		2	1	4
		FULLY SUS			2	3	5
		10667			1	1	2
		COT			1	1	2
		10800		1			1
		FULLY SUS		1			1
		10853			1		1
		FULLY SUS			1		1
		10862			1		1
		FULLY SUS			1		1
		10950				1	1
		COT				1	1
		11055				1	1
		FULLY SUS				1	1
		11072				1	1
		FULLY SUS				1	1
		11090			1		1
		PEN_COT			1		1
		11111			1		1
		FULLY SUS			1		1
		11772	1	1	1		3
		FULLY SUS	1	1	1		3
		12137			1		1
		FULLY SUS			1		1
		12550	1				1

		COT	1				1
		12551		1			1
		COT		1			1
		17796				1	1
		COT				1	1
	GPSC46	30			1	1	2
		COT			1	1	2
		3450	1		8	14	23
		COT	1		8	8	17
		FULLY SUS				6	6
		10584	3	1		3	7
		COT	1	1		3	5
		FULLY SUS	2				2
		10673			2	6	8
		FULLY SUS			2	6	8
		10733	1				1
		COT	1				1
		10858			1	3	4
		FULLY SUS			1	3	4
		17617				1	1
		COT				1	1
		17634				1	1
		FULLY SUS				1	1
		17744				1	1
		FULLY SUS				1	1
	GPSC114	5326		2	7	8	17
		COT		2	6	6	14
		FULLY SUS			1	2	3
		8827	1				1
		FULLY SUS	1				1
		10787	1				1
		COT	1				1
		12530	1				1
		COT	1				1
		17773				1	1
		COT				1	1
		17778				1	1
		FULLY SUS				1	1
	GPSC207	9815			1	1	2
		COT			1	1	2
		11723			1	1	2
		COT			1	1	2

	GPSC428	10730	1				1
		FULLY SUS	1				1
19A	GPSC17	2057				3	3
		PEN_COT				3	3
		2062	50	38	66	87	241
		PEN	3				3
		PEN MDR				1	1
		PEN_COT	47	38	66	86	237
		4869			1		1
		PEN_COT			1		1
		4872		1	1	1	3
		PEN_COT		1	1	1	3
		4873			1		1
		OTHER RESISTANCE			1		1
		6121	5	3	11	2	21
		PEN_COT	5	3	11	2	21
		6255		1		1	2
		PEN_COT		1		1	2
		6258	1		2	2	5
		PEN_COT	1		2	2	5
		6261			1		1
		PEN_COT			1		1
		6262			2		2
		PEN_COT			2		2
		6266	2		1		3
		PEN_COT	2		1		3
		6268	1				1
		PEN_COT	1				1
		7895		1	1	1	3
		PEN MDR		1			1
		PEN_COT			1	1	2
		7899			1		1
		PEN_COT			1		1
		7903	3				3
		PEN MDR	2				2
		PEN_COT	1				1
		8309				1	1
		PEN_COT				1	1
		8659				1	1
		PEN				1	1
		9900			1		1
		PEN_COT			1		1

		10690	1				1
		PEN_COT	1				1
		10693	1				1
		PEN_COT	1				1
		10741	1				1
		PEN	1				1
		10756		1			1
		PEN_COT		1			1
		10790	1				1
		PEN_COT	1				1
		10843			1		1
		PEN_COT			1		1
		10870			1		1
		PEN_COT			1		1
		11085				1	1
		PEN_COT				1	1
		11122			1		1
		PEN_COT			1		1
		11677	1		1		2
		PEN_COT	1		1		2
		12226			1		1
		PEN_COT			1		1
		12521			1		1
		PEN_COT			1		1
		12564			1		1
		PEN_COT			1		1
		14429	1				1
		PEN_COT	1				1
		17030				1	1
		PEN_COT				1	1
	GPSC5	172	4	2	1		7
		PEN	3	1	1		5
		PEN MDR		1			1
		PEN_COT	1				1
		2218		1			1
		PEN_COT		1			1
		9913	2		2		4
		PEN MDR	2		1		3
		PEN_COT			1		1
		10744		1			1
		PEN_COT		1			1
	GPSC10	276				3	3

		PEN MDR				3	3
		2013				1	1
		PEN MDR				1	1
		5033	1				1
		PEN MDR	1				1
	GPSC14	2059				2	2
		PEN MDR				2	2
		8299			1		1
		PEN MDR			1		1
		8741		1			1
		PEN MDR		1			1
	GPSC1	320			1	1	2
		PEN MDR			1	1	2
		12586			1		1
		PEN MDR			1		1
	GPSC4	199	1				1
		FULLY SUS	1				1
		416			1		1
		FULLY SUS			1		1
	GPSC33	4088		1			1
		FULLY SUS		1			1
	GPSC90	10703	1				1
		PEN_COT	1				1
	GPSC21	6095	1				1
		PEN_COT	1				1
19F	GPSC1	271	21	7	13	13	54
		PEN MDR	21	7	13	13	54
		420	7	9	8	3	27
		PEN MDR	1		1		2
		PEN_COT	6	9	7	3	25
		1412	1				1
		PEN MDR	1				1
		2706	1				1
		PEN MDR	1				1
		2834	1				1
		PEN MDR	1				1
		2920	1				1
		PEN MDR	1				1
		8678			3	7	10
		PEN MDR			3	7	10
		9933			1		1
		PEN_COT			1		1

		10683	1				1
		COT	1				1
		10731	1				1
		PEN_COT	1				1
		10822		1			1
		PEN MDR		1			1
		10825		1			1
		PEN MDR		1			1
		11129			1		1
		PEN MDR			1		1
		11696		1	1	2	4
		PEN MDR		1	1	2	4
		12134			1		1
		PEN MDR			1		1
		12136			1	1	2
		PEN MDR			1	1	2
		12228	1				1
		PEN MDR	1				1
		12229	1				1
		PEN MDR	1				1
		12526	1				1
		PEN MDR	1				1
		12533	1				1
		PEN MDR	1				1
		12534	1				1
		PEN MDR	1				1
		12536	1				1
		PEN MDR	1				1
		12537	1				1
		PEN MDR	1				1
		12543	1				1
		PEN_COT	1				1
		12560	1				1
		PEN MDR	1				1
	GPSC21	347	16	10	12	24	62
		PEN MDR	1	1		1	3
		PEN_COT	15	9	12	23	59
		5179	2				2
		PEN_COT	2				2
		6277	4	3	4	2	13
		PEN_COT	4	3	4	2	13
		8670				1	1

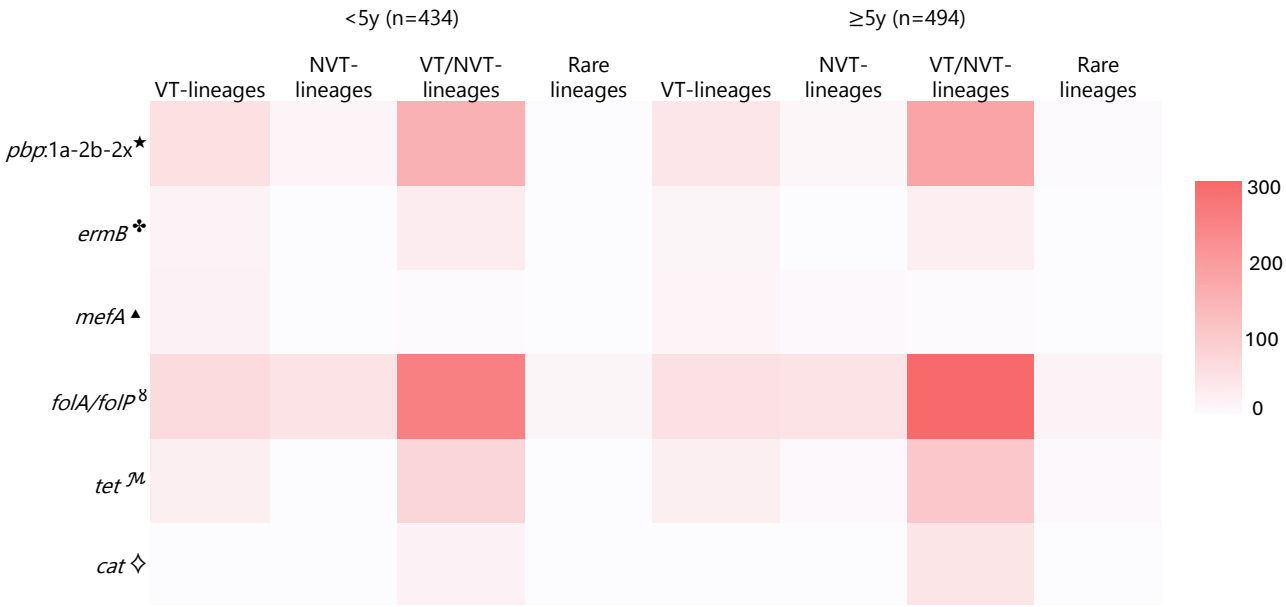
		COT				1	1
		9927			1		1
		PEN_COT			1		1
		10536	1				1
		PEN_COT	1				1
		10553	2				2
		PEN_COT	2				2
		10572	1				1
		PEN_COT	1				1
		10619	1	1			2
		PEN_COT	1	1			2
		10709	1				1
		PEN_COT	1				1
		10752		1			1
		PEN_COT		1			1
		10784	1				1
		FULLY SUS	1				1
		10794		1			1
		PEN_COT		1			1
		10798		1			1
		PEN_COT		1			1
		10935				1	1
		PEN_COT				1	1
		12146			1		1
		PEN_COT			1		1
		12585			1		1
		PEN_COT			1		1
		12700			1		1
		PEN_COT			1		1
		16974				1	1
		PEN_COT				1	1
		17014				2	2
		PEN_COT				2	2
		17033				1	1
		PEN_COT				1	1
		17155				2	2
		PEN_COT				2	2
	GPSC90	8326	2	2		2	6
		PEN_COT	2	2		2	6
		8328	3	1		1	5
		PEN_COT	3	1		1	5
		8332	1				1

		PEN_COT	1				1
		9928			1		1
		PEN MDR			1		1
		10827		1			1
		PEN_COT		1			1
		11681	2				2
		PEN_COT	2				2
		12145			1		1
		PEN_COT			1		1
		12522			1		1
		PEN_COT			1		1
	GPSC205	2067	5	1	1		7
		PEN_COT	5	1	1		7
		10716	1				1
		PEN_COT	1				1
		10823		1		5	6
		PEN_COT		1		5	6
		12535	1			1	2
		PEN_COT	1			1	2
		17024				1	1
		PEN_COT				1	1
	GPSC5	172		2	3	5	10
		PEN_COT		2	3	5	10
		10817		1			1
		PEN_COT		1			1
	GPSC159	7131	2		1		3
		PEN_COT	2		1		3
		9937		1	1	1	3
		PEN_COT		1	1	1	3
		10706	1				1
		PEN MDR	1				1
		10795		1			1
		PEN_COT		1			1
	GPSC17	2062	2				2
		PEN_COT	2				2
		7897	1				1
		PEN_COT	1				1
	GPSC192	10740		1			1
		FULLY SUS		1			1
		10762		1			1
		COT		1			1
	GPSC44	179				1	1

		PEN MDR				1	1
		13405				1	1
		OTHER RESISTANCE				1	1
	GPSC48	1203	2				2
		PEN	2				2
	GPSC725	12527	1				1
		PEN_COT	1				1
	GPSC198	87	1				1
		PEN MDR	1				1
23B	GPSC102	4423			2	13	15
		PEN MDR				1	1
		PEN_COT			2	12	14
		7687			2	4	6
		COT			2	4	6
		17031				1	1
		COT				1	1
		17558				1	1
		COT				1	1
	GPSC5	1349				5	5
		PEN_COT				5	5
		14492				1	1
		PEN				1	1
	GPSC169	2911	2	1	2		5
		COT	1		1		2
		FULLY SUS	1	1	1		3
	GPSC426	10719	1				1
		FULLY SUS	1				1
	GPSC7	355				1	1
		FULLY SUS				1	1
	GPSC41	1094	1				1
		PEN_COT	1				1
35B	GPSC5	171				1	1
		PEN_COT				1	1
		172			1	4	5
		PEN_COT			1	4	5
		361		1	25	38	64
		PEN_COT		1	25	38	64
		12126			1		1
		PEN_COT			1		1
		12580			1	1	2
		PEN_COT			1	1	2
		17718				1	1

		PEN_COT				1	1
		17919				1	1
		PEN_COT				1	1
	GPSC160	9813	2		5	2	9
		COT	2		3	2	7
		FULLY SUS			2		2
	GPSC233	5396			1	4	5
		COT			1	2	3
		FULLY SUS				2	2
	GPSC30	8774		1	1	1	3
		COT		1	1	1	3
	GPSC255	17022				1	1
		COT				1	1
		17785				1	1
		FULLY SUS				1	1
	GPSC59	558				2	2
		OTHER RESISTANCE				2	2
	GPSC72	3329				1	1
		COT				1	1
Total			**242**	**125**	**366**	**514**	**1247**

Appendix R



Antimicrobial resistance by lineage type in the late-PCV13 period (2015–2020), stratified by age category (<5 and ≥5 years, n=928). The red gradient colour scheme is proportional to the number of resistance determinants, with dark red correlating to higher frequencies.

★ Penicillin-binding protein (PBP) combinations for penicillin resistance prediction

✣ Erythromycin resistance methylase gene conferring resistance to erythromycin and clindamycin

▲ Macrolide efflux pump gene conferring resistance to erythromycin

⁸ Cotrimoxazole resistance conferred by the presence of mutation I100L in *folA* and/or indel within amino acid residue 56-67 in *folP*

ℳ Tetracycline resistance genes (*tet*M or *tet*S/M) conferring resistance to tetracycline

✧ Chloramphenicol acetyltransferase gene conferring chloramphenicol resistance

Appendix S

Serotype 8 isolates by country, vaccine period, source of specimen, and age group.

Country	Vaccine introduction year	n	Vaccine period		Source of specimen		Age category	
			Pre-PCV n (%)	Post-PCV n (%)	Disease n (%)	Carriage n (%)	<5 years n (%)	≥5 years n (%)
Bangladesh	2015 (PCV10)	1	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)
Belarus	2014 (PCV10)	2	0 (0)	2 (100)	0 (0)	2 (100)	1 (50)	1 (50)
Brazil	2010 (PCV10)	10	2 (20)	8 (80)	10 (100)	0 (0)	4 (40)	6 (60)
Bulgaria*†	No PCV‡	1	_	_	1 (100)	0 (0)	_	_
China	2009 (PCV7); 2010 (PCV10); 2011 (PCV13)	1	0 (0)	1 (100)	0 (0)	1 (100)	1 (100)	0 (0)
Egypt*	No PCV‡	1	_	_	1 (100)	0 (0)	1 (100)	0 (0)
Ethiopia*†	No PCV‡	1	_	_	0 (0)	1 (100)	_	_
France†	2006 (PCV7); 2010 (PCV13)	1	1 (100)	0 (0)	1 (100)	0 (0)	_	_
India◇✧†	2017 (PCV13)	10	5 (50)	4 (40)	9 (90)	_	1 (10)	8 (80)
Israel	2009 (PCV7); 2010 (PCV13)	11	3 (27)	8 (73)	11 (100)	0 (0)	8 (73)	3 (27)
Malawi†	2011 (PCV13)	5	2 (40)	3 (60)	4 (80)	1 (20)	3 (60)	1 (20)
Malaysia*	No PCV‡	2	_	_	2 (100)	0 (0)	0 (0)	2 (100)
Mongolia*†	No PCV‡	1	_	_	1 (100)	0 (0)	_	_
Morocco	2010 (PCV13); 2012 (PCV10)	1	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)
Mozambique†	2013 (PCV10)	3	3 (100)	0 (0)	3 (100)	0 (0)	2 (67)	1 (33)
Nepal	2015 (PCV10)	13	8 (62)	5 (39)	3 (23)	10 (77)	11 (85)	2 (15)
Netherlands✧	2011 (PCV10)	178	43 (24)	135 (76)	_	_	5 (3)	173 (97)
New Zealand	2008 (PCV7); 2011 (PCV10); 2014 (PCV13)	5	0 (0)	5 (100)	5 (100)	0 (0)	5 (100)	0 (0)
Nigeria	2014 (PCV10)	1	0 (0)	1 (100)	1 (100)	0 (0)	1 (100)	0 (0)
Papua New Guinea*	No PCV‡	6	_	_	6 (100)	0 (0)	6 (100)	0 (0)
Poland*	No PCV‡	2	_	_	2 (100)	0 (0)	2 (100)	0 (0)
Russian Federation	2014 (PCV13)	5	5 (100)	0 (0)	5 (100)	0 (0)	0 (0)	5 (100)
Slovenia*	No PCV‡	2	_	_	2 (100)	0 (0)	2 (100)	0 (0)
South Africa	2009 (PCV7); 2011 (PCV13)	391	30 (8)	361 (92)	391 (100)	0 (0)	218 (56)	173 (44)
Thailand ▲	No PCV‡	2	_	_	2 (100)	0 (0)	0 (0)	2 (100)
The Gambia*	2009 (PCV7); 2011 (PCV13)	10	0 (0)	10 (100)	1 (10)	9 (90)	3 (30)	7 (70)
Turkey	2008 (PCV7)	1	0 (0)	1 (100)	1 (100)	0 (0)	0 (0)	1 (100)
United States	2000 (PCV7); 2010 (PCV13)	19	2 (11)	17 (90)	19 (100)	0 (0)	7 (37)	12 (63)

‡ No vaccine introduction

*Vaccine period data not included

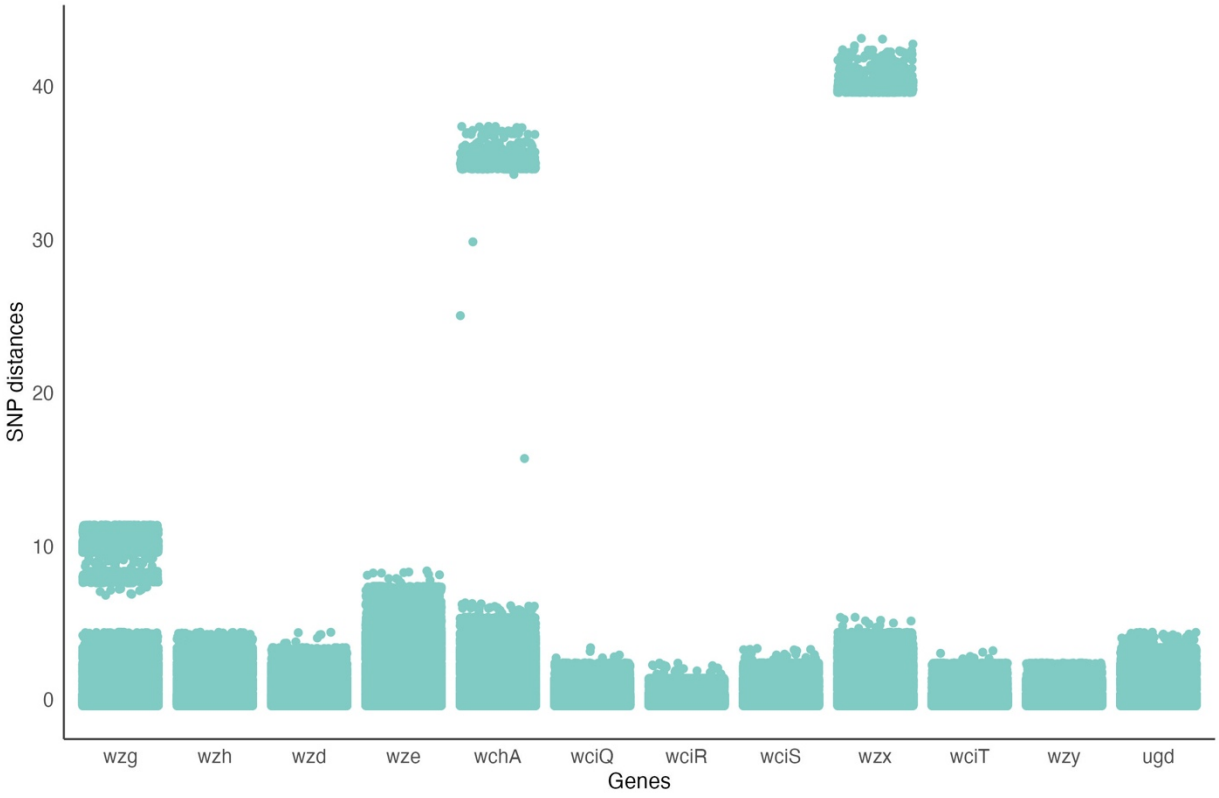
◇Some or all isolates missing year information

†Some or all isolates missing age information

✧ Some or all isolates missing source information

▲ PCV10 and PCV13 are available in Thailand, but not included in the National Immunization Programme

Appendix T



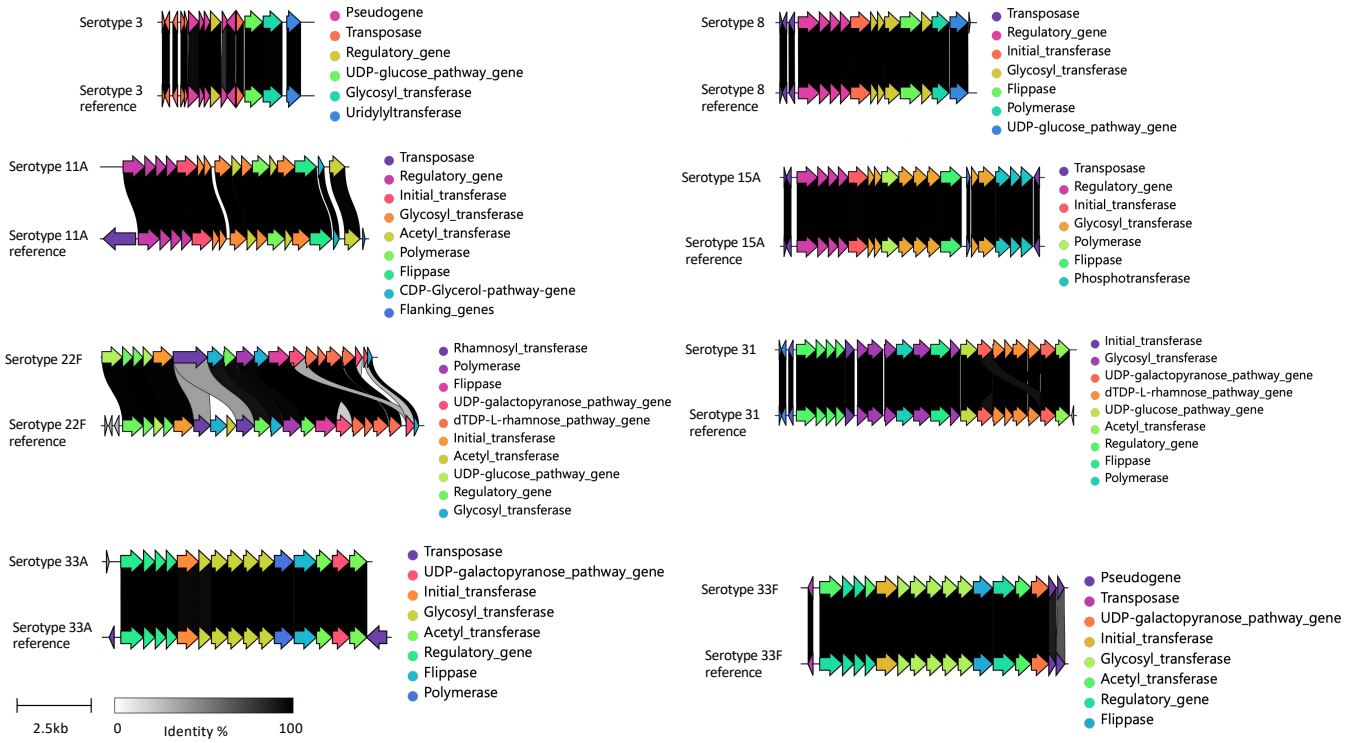
Serotype 8 capsular locus (CPS) single nucleotide polymorphism (SNP) analysis, N=686. SNP variations by gene from the serotype 8 reference sequence (CR931644) are expressed by SNP distances.

Appendix U

Distribution of lineages expressing serotype 8 by continent and country of origin.

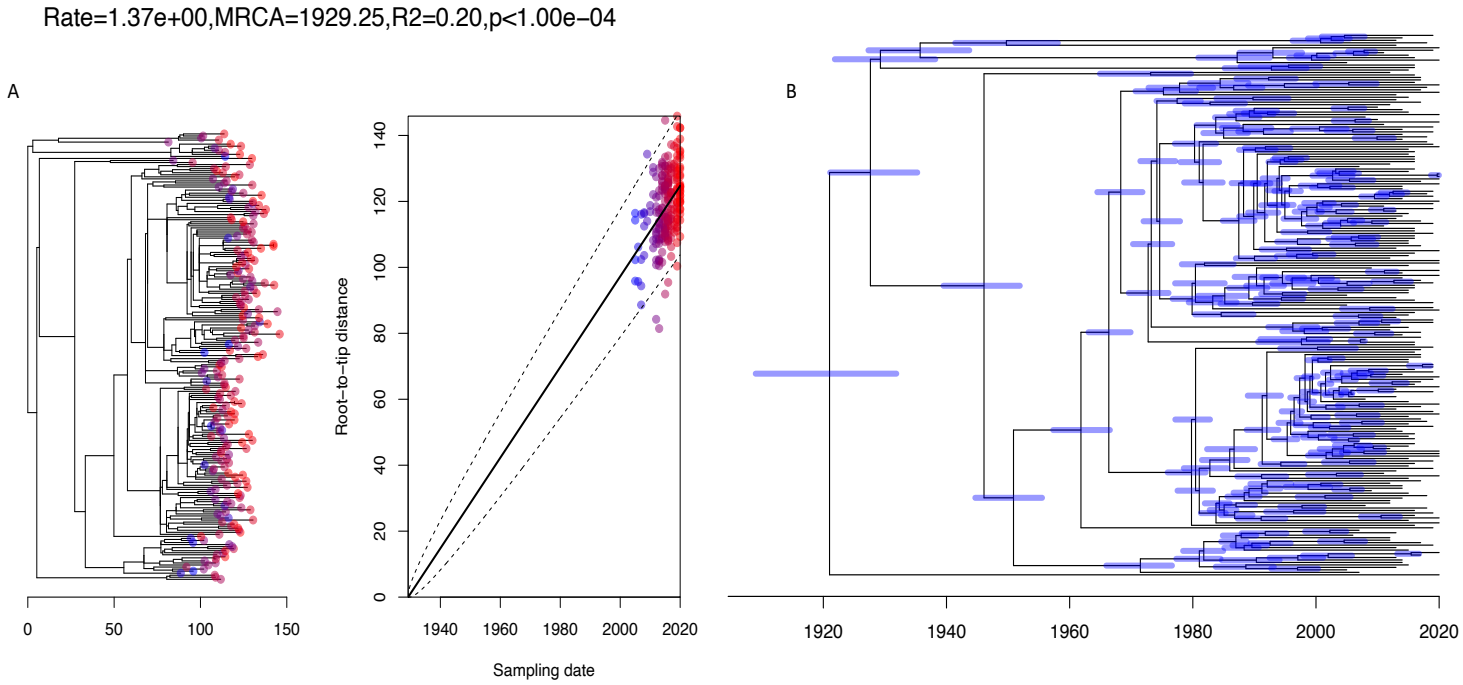
Continent	GPSC lineages									
Country	3 (n=581)	22 (n=1)	32 (n=4)	98 (n=47)	224 (n=34)	277 (n=10)	336 (n=6)	694 (n=2)	732 (n=1)	Total (n=686)
Africa	376	1	-	22	9	5	-	-	-	413
Egypt	-	-	-	1	-	-	-	-	-	1
Ethiopia	-	-	-	-	1	-	-	-	-	1
Malawi	-	-	-	1	-	4	-	-	-	5
Morocco	1	-	-	-	-	-	-	-	-	1
Mozambique	2	-	-	1	-	-	-	-	-	3
Nigeria	-	-	-	1	-	-	-	-	-	1
South Africa	373	1	-	10	6	1	-	-	-	391
The Gambia	-	-	-	8	2	-	-	-	-	10
Asia	-	-	-	-	17	5	6	1	1	30
Bangladesh	-	-	-	-	1	-	-	-	-	1
China	-	-	-	-	1	-	-	-	-	1
India	-	-	-	-	-	4	5	-	1	10
Malaysia	-	-	-	-	1	-	-	1	-	2
Mongolia	-	-	-	-	1	-	-	-	-	1
Nepal	-	-	-	-	11	1	1	-	-	13
Thailand	-	-	-	-	2	-	-	-	-	2
Europe	188	-	4	10	-	-	-	1	-	203
Belarus	1	-	-	1	-	-	-	-	-	2
Bulgaria	1	-	-	-	-	-	-	-	-	1
France	1	-	-	-	-	-	-	-	-	1
Israel	9	-	-	2	-	-	-	-	-	11
Netherlands	169	-	1	7	-	-	-	1	-	178
Poland	2	-	-	-	-	-	-	-	-	2
Russian Federation	2	-	3	-	-	-	-	-	-	5
Slovenia	2	-	-	-	-	-	-	-	-	2
Turkey	1	-	-	-	-	-	-	-	-	1
Latin America	9	-	-	1	-	-	-	-	-	10
Brazil	9	-	-	1	-	-	-	-	-	10
North America	3	-	-	14	2	-	-	-	-	19
United States	3	-	-	14	2	-	-	-	-	19
Oceania	5	-	-	-	6	-	-	-	-	11
New Zealand	5	-	-	-	-	-	-	-	-	5
Papua New Guinea	-	-	-	-	6	-	-	-	-	6

Appendix V



Annotated gene clusters of capsular loci (CPS) of serotypes expressed by GPSC3. Each serotype is aligned against its reference.

Appendix W



Clonal complex 53 (CC53) Bayesian inference of ancestral dates using BactDating among South African isolates (N=219), 2005–2020. A) Root-to-tip distance regression analysis for temporal signal detection of recombination-free CC53 phylogeny. Tips are shaded in a colour gradient from red (more recent isolates) to blue (older isolates). B) Time-resolved phylogeny showing ancestral dates with 95% credible intervals represented by blue shades.