Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Tyler Landrith _____ April 20, 2012
Date

Descriptive Epidemiology and Analysis of Risk Factors for Invasive Disease Due to Serotype 1 *Streptococcus pneumoniae* in South Africa from 2003-2008

By

Tyler Landrith
MPH

Epidemiology

_____

Dr. Keith Klugman, MD, PhD
Committee Chair

_____

Dr. Anne Von Gottberg, MD
Committee Chair

Descriptive Epidemiology and Analysis of Risk Factors for Invasive Disease Due to Serotype 1 *Streptococcus pneumoniae* in South Africa from 2003-2008

By
Tyler Landrith

Bachelor of Science
University of California, Davis
2009

Thesis Committee Chairs:
Dr. Keith Klugman, PhD, William A. Foege Professor of Global Health,
Emory University, Atlanta, GA, USA

Dr. Anne von Gottberg, MD, Head of Unit, RMPRU,
NICD, Johannesburg, South Africa

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2012

Abstract

Descriptive Epidemiology and Analysis of Risk Factors for Invasive Disease Due to Serotype 1 *Streptococcus pneumoniae* in South Africa from 2003-2008

By Tyler Landrith

Introduction:

Streptococcus pneumoniae is responsible for a variety of human infections, and is a significant cause of morbidity and mortality among infants and the elderly, particularly in developing countries. The pneumococcus has a total of 93 known immunologically distinct polysaccharide capsules known as serotypes. Each serotype, including serotype 1 can exhibit unique epidemiological patterns based on factors such as invasiveness, age, outbreak potential, disease outcome, and HIV infection. This study sought to elucidate these patterns during the period of interest using three different epidemiological methods.

Methods:

A description of cases and putative outbreaks of serotype 1 was conducted according to time and place during the period 2003-2008 using data from a laboratory-based surveillance system. Bivariate case-control analyses were used to test for significant associations between serotype 1 and all relevant literature-based hypothesized exposures, including HIV infection, relative to other serotypes causing invasive pneumococcal disease. Bivariate analyses were also used to test for significant associations between serotype 1 IPD and clinically relevant outcomes. A multivariable model was created for refined risk factors for serotype 1 IPD using logistic regression. Several multivariable models were created to test the significance of serotype 1 IPD as a predictor of clinical outcomes, controlling for relevant covariates.

Results:

Serotype 1 was the most frequently isolated type in reported cases of IPD over the period of interest (n=2701). The seven most frequently isolated serotypes over the period of interest are all contained in the PCV13 vaccine, which is currently being rolled out in South Africa. In Gauteng province, putative outbreaks occurred during July 2003 and October 2008. Gauteng province also exhibited apparent seasonal variation, with June and July having the highest reported number of cases across time. Multivariable analysis revealed an age category of 5-17 and combined smoking/alcohol behaviors to be specific risk factors for serotype 1 IPD. Serotype 1 IPD was a significant predictor of a clinical diagnosis of meningitis ($p<0.01$) and LRTI ($p=0.04$) at discharge when controlling for HIV infection at time of admission, underlying conditions other than HIV, antibiotic use, age, race, and province.

Descriptive Epidemiology and Analysis of Risk Factors for Invasive Disease Due to Serotype 1 *Streptococcus pneumoniae* in South Africa from 2003-2008

By
Tyler Landrith

Bachelor of Science
University of California, Davis
2009

Thesis Committee Chairs:

Dr. Keith Klugman, MD, PhD, William H. Foege Professor of Global Health, Emory University, Atlanta, GA, USA

Dr. Anne von Gottberg, MD, Head of Unit, RMPRU, NICD, Johannesburg, South Africa

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2012

Acknowledgements:

I am grateful first and foremost to my advisors, Dr. Keith Klugman and Dr. Anne von Gottberg. I owe both equal thanks for their numerous tireless insights, inputs, and assistance at all steps of this process. Without them, I would not have had the opportunity to work with this data in South Africa, and this thesis would not exist. I would like to thank Dr. Klugman for identifying this project as one that would be of interest to me, and continuing to support my academic growth. I would like to thank Dr. von Gottberg for providing her data and hosting me at the Respiratory and Meningeal Pathogens Reference Unit. Thank you for not only advising me, but also doing so much to help me make the most of my time in South Africa.

I would like to acknowledge the Global Field Experience committee, the Eugene J. Gangarosa Fund, the William A. Foege Global Health Fund, and the O.C. Hubert Fellowships for International Health for providing the funding for this research.

I would like to give a heartfelt thanks to the South African patients who consented to provide the clinical data that was used in this thesis. I hope that this research will both honor and benefit them.

Thanks to Linda de Gouveia for extracting and compiling the raw dataset for analysis from Access, you made my job much easier. Thanks also for your availability to answer questions and for all you did to welcome me to SA, and adjust to life there.

Thank you to Phyllis Hyde for your invaluable role in helping me to feel at home during my stay at the NICD.

Thanks to Catherine Bozio for reviewing my initial drafts and providing valuable suggestions for my analysis plan.

Thank you to Kelly McCormick for the biostatistical support, particularly with the multivariable portion of the analysis.

To Emily Dloboyi, Judith Tshabalala, Portia Mogale, and Thembe Mthembu: Thank you for answering my many questions about the dataset. Thank you also for allowing me to share a workspace with you for two months and making me feel welcome in so many ways. I will always remember to answer “unjani?” with “mjojo!” (please forgive my spelling).

To all the RMPRU staff and students: Thank you for all your support and for helping me to feel welcome during my stay in your Unit. You are each an inspiration to me.

I would like to also thank the faculty and students at the Rollins School of Public Health; you have been an invaluable source of inspiration and support during my time here.

Thanks to my family and friends, to my parents Mark and Michele, and my brothers Aaron and Garrett for supporting my goal to pursue this degree in so many ways. I knew I could always turn to you for encouragement and support, no matter what.

Finally, I would like to give a special thanks to Kindra Eidson. Thank you for your consistent sacrifice, encouragement, and support while I pursued this degree. There is absolutely no way I could have done this without you. I can say with all seriousness that you earned this degree just as much as I did.

Table of Contents

I. PROJECT OVERVIEW: 10

II. INTRODUCTION: 12

1. JUSTIFICATION FOR STUDY: 12

2. LITERATURE REVIEW: 12

2.1 Serotype-Specific Epidemiology - Invasiveness 12

2.2 Antibiotic Resistance 13

2.3 Clinical Syndrome 14

2.4 Vaccines 15

2.5 Age Specific Factors 17

2.6 HIV 18

2.7 Outbreaks 19

2.8 Anti-Retroviral Therapy and Cotrimoxazole prophylaxis 20

2.9 Relationship Between Serotype and Clinical Outcomes, including Mortality and Severity of Disease 22

3. INTENDED/POTENTIAL USE OF STUDY FINDINGS: 23

4. STUDY DESIGN/LOCATION: 23

5. OBJECTIVES: 24

6. HYPOTHESES: 25

III. PROCEDURES/METHODS: 28

1. STUDY POPULATION: 28

1.1 Description and source of study population 28

1.2 Case Definition 29

1.3 Inclusion/exclusion criteria 29

1.4 Sampling, including sample size and statistical power 29

1.5 Consent process 30

1.6 Audience and stakeholder participation 30

2. VARIABLES: 31

2.1 Surveillance and Laboratory Methods 34

3. DATA HANDLING AND ANALYSIS: 35

3.1 Analysis 35

3.2 Quality Control/Assurance 41

3.3 Faculty and Ethical Approval 41

3.4 Handling of Adverse Events: 41

IV RESULTS: 43

1. DESCRIPTIVE ANALYSIS 43

1.1 Basic epidemiological characteristics 43

1.2 Seasonality 43

2. BIVARIATE ANALYSIS 44

2.1 Hypothesized unique risk factors for serotype 1 IPD 44

2.2 Antiretroviral therapy and cotrimoxazole prophylaxis 44

2.3 Clinical syndrome, disease severity, and mortality 45

3. MULTIVARIABLE ANALYSIS 45

3.1 Initial Model for Serotype 1 IPD Risk Factors 45

3.2 Confounding Assessment and Final Model 46

3.3 Analysis of antiretroviral use among HIV positive patients with IPD infection 46
3.4 Multivariable models containing serotype 1 IPD as a predictor of clinical outcomes . 47

V DISCUSSION: 49
1. STRENGTHS AND WEAKNESSES 49
2. DISCUSSION OF FINDINGS 50

VI CONCLUSION: 56

VII REFERENCES: 57

VIII. FIGURES AND TABLES 65
FIGURE 1 65
FIGURE 2 65
FIGURE 3 66
FIGURE 4 66
FIGURE 5 67
FIGURE 6 67
FIGURE 7 68
FIGURE 8 72
TABLE 1 75
TABLE 2 76
TABLE 3 76
TABLE 4 77
TABLE 5 77
TABLE 6 78
TABLE 7 78
TABLE 8 79
TABLE 9 79
TABLE 10 80

I. Project Overview:

Streptococcus pneumoniae is responsible for a variety of human infections, and is a significant cause of morbidity and mortality among infants and the elderly, particularly in developing countries. The pneumococcus has a total of 93 known immunologically distinct polysaccharide capsules known as serotypes. Each serotype, including serotype 1, follows unique epidemiological patterns based on factors such as invasiveness, age, outbreak potential, and disease outcome (1). Of particular interest in this study is the relationship between HIV infection and invasive pneumococcal disease due to serotype 1. Though the type is less often isolated from HIV infected patients than other types the condition has not been used as a predictor of serotype 1 IPD (2,3). This study seeks to describe and analyze the epidemiology of serotype 1 in South Africa through three primary objectives. These are the following:

1. A description of cases and putative outbreaks of serotype 1 according to time and place during the period 2003-2008 using data from a laboratory-based surveillance system.
2. Perform bivariate analyses for hypothesized exposures and covariates.
    a. Use of bivariate case-control analyses to test for significant associations between serotype 1 and all relevant literature-based hypothesized exposures, including HIV infection, relative to other serotypes causing invasive pneumococcal disease.
    b. Use of bivariate analyses to test for significant associations between serotype 1 IPD and clinically relevant outcomes.

3. Perform multivariable analysis for hypothesized exposures and covariates
    a. Creation of a model for refined risk factors for serotype 1 IPD using a multivariable logistic regression.
    b. Creation of a multivariable model to test the significance of serotype 1 IPD as a predictor of clinically relevant outcomes, controlling for significant covariates.

II. Introduction:

1. Justification for study:

Streptococcus pneumoniae continues to be responsible for cases of acute respiratory infection, otitis media, meningitis, and bacteraemia/septicaemia worldwide (1-6). In 2010, the National Institute for Communicable Diseases (NICD) in South Africa reported 4205 laboratory-confirmed cases of invasive pneumococcal disease (IPD) among all age groups throughout South Africa (7). Of these cases, 907 were reported to be in children less than 5 years of age (7). As cases are drawn from a national surveillance system, these numbers only represent a small proportion of the true disease burden in the country, but are still useful for describing outbreaks and risk factors unique to serotype 1 (8). There has not been a study using an active population-based surveillance system to describe outbreaks of serotype 1 *S. pneumoniae* in South Africa during the time period of interest while developing a multivariable logistic regression model to extract unique potential risk factors for serotype 1 cases in a pre-vaccine setting.

2. Literature review:

2.1 Serotype-Specific Epidemiology - Invasiveness

A polysaccharide capsule surrounds the cell wall of *S. pneumoniae*, serving in part as a mechanism for the bacterium to colonize the upper respiratory tract and evade the host immune system (9-11). There are 93 known immunologically distinct

polysaccharide capsules, or serotypes, within the species (1, 4). There is evidence that each serotype or serogroup of *S. pneumoniae* is epidemiologically unique, particularly in terms of invasiveness and frequency of isolation from the human nasopharynx (1, 3). Serotype 1 is characterized as being highly invasive, but rarely isolated from carriage (1). According to Brueggemann et al, invasiveness can be considered as an index comparing the frequency of asymptomatic carriage versus the frequency of isolation from a normally sterile site, usually blood or cerebrospinal fluid (CSF) (6). Thus serotype 1 has a high invasiveness index in the sense that it is often isolated from a normally sterile site, but is presumed to be carried for only a brief period of time (6). Other serotypes that have been reported to be highly invasive include types 4, 5, 14, and 18C. (1,6).

2.2 Antibiotic Resistance

Streptococcus pneumoniae is a highly adaptable pathogen, and particularly as a result of antibiotic overuse, many strains have developed resistance to treatment (12-14). Serotype 1 usually remains susceptible to antibiotics, particularly of the β-lactam class, especially when the antibiotic is given according to the correct dosing regimen (1,15).

Hausdorff et al. suggest that serotypes commonly isolated from carriage in the nasopharynx are more likely to develop resistance to antibiotics, and this may explain why serotype 1, along with other "non-carriage" serotypes, remains susceptible (1). There is good observational evidence from the literature to support this. In particular, a 2009 study by Fenoll et al. examined temporal trends in serotype prevalence, antibiotic resistance, and IPD in Spain during the time period from 1979 to 2007 (16). Despite a significant overall increase from 5% of IPD isolates in 1997 to 14% in 2007, there was no

corresponding increase in resistance to either penicillin or erythromycin, with the percentage of resistant serotype 1 isolates remaining well below 10% (16). In contrast, the percentage of antibiotic resistant carriage serotypes increased in parallel with increased percentage of IPD isolates (16). This relationship between serotype prevalence across time and antibiotic resistance also has important implications for vaccine efficacy (see Section 2.4)

2.3 Clinical Syndrome

In a clinical setting as well as in the literature, a case of IPD is usually regarded as isolation from a normally sterile site, which includes CSF or blood. This excludes, by definition, contiguous spread from the nasopharynx as a mode of infection, though it is possible for contiguous spread to occur in cases of otitis media or sinusitis and then to complicate and cause meningitis (11). Very rarely, occult bacteremia can serve as a clinical indication that the pneumococcus has become invasive (11). Once the pneumococcus has entered the bloodstream, it can seed to an infection site, causing various presentations, including but not limited to pneumonia, meningitis, peritonitis or endocarditis (11). Occult bacteremia occurs when the bacterium does not or has not yet reached a specific infection site, other than the blood. According to Hausdorff et al., clinical syndromes are closely related to other epidemiological factors of IPD due to serotype 1, including age and hospitalization rate (1). To this point, serotype 1 is more commonly isolated in cases of hospitalization due to bacteremia, particularly among children younger than 6 (1). This contrasts with the fact that young children tend to present with pneumococcal meningitis more often than older children (1).

2.4 Vaccines

The first pneumococcal vaccine, licensed by Merck, contained 23 free capsular polysaccharides, which includes serotype 1. Currently, the 23 polysaccharide vaccine is recommended for use in adults and older children, but does not induce an immune response in children younger than two (9, 17-18). The pneumococcal conjugate vaccine 7 (PCV7), licensed by Pfizer, contains 7 *S. pneumoniae* capsular polysaccharides conjugated to a carrier protein, which is recognized as T-cell dependent, and subsequently stimulates a humoral immune response with T – cell help (11). The serotypes included in this vaccine formulation are 4, 6B, 9V, 14, 18C, 19F, and 23F (1). These vaccines do not include serotype 1, but contain serotypes that are responsible for most pediatric IPD, notably in the United States (1). In South Africa, PCV7 entered the private sector in 2005, but did not become the standard of care until 2009. WHO recommends the vaccine to be given to infants as a 3 part series at 6, 10, and 14 weeks in developing countries (16). In South Africa, the vaccine is given at 6 and 14 weeks, with a booster at 9 months according to recommendations by the Expanded Programme on Immunisation (EPI) (18).

As serotype 1 (along with serotype 5) is a significant cause of IPD in developing countries, randomized, double-blind, placebo-controlled clinical trials (RCTs) of a 9-valent vaccine including these serotypes were conducted among children in South Africa and the Gambia (19-20). The trials, although successful in reducing mortality and hospital visits due to pneumonia in the Gambia, and in reducing incidence of IPD in South Africa, failed to confirm protection against serotype 1 (19-21). This was perhaps based on failure after 18 months of age in protection against the serotype, when the vaccine was given without a booster (21).

Currently, the PCV13 vaccine is replacing PCV7 in South Africa (22). As with PCV7, the vaccine is given as a 2+1 schedule (two injections plus one booster) at 6 weeks, 14 weeks, and 9 months in the public sector. In addition to the serotypes contained in the 7 valent vaccine, the 13-valent contains serotypes 1, 5, 7F, 19A, 3, and 6A.

Nunes et al. and Klugman both suggest the risk of serotype replacement as a result of vaccination with PCV7, noting that this could increase the frequency of carriage for non-PCV7 serotypes (23-24). One such study providing strong evidence that this does occur was a double-blind RCT of a 9-valent conjugate vaccine on 500 infants in Soweto conducted by Mbelle et al. (25). The effect of the vaccine on pneumococcal carriage was examined as part of their study. The authors report a significant ($p=0.007$) increase in carriage of non-vaccine serotypes from 25% to 36% of vaccinees, which corresponded to a significant ($p<0.001$) decrease in carriage of vaccine (24). Additionally, Cutts et al. reported a non-significant increase in non-PCV7 IPD in the Gambia RCT, which provides indirect evidence that this phenomenon does occur (19).

In this vein, a 2011 review article by Weinberger et al states that although strong evidence exists for replacement in carriage, there is not a corresponding effect in disease caused by the replacement serotypes (26). They suggest that this may be a result of a combination of factors: lower invasiveness of replacing serotypes, biases in carriage data prior to the vaccine, and/or underestimation of replacement by surveillance systems.

Nonetheless, there is more recent evidence that incomplete disease replacement does occur. Scott et al. conducted a study comparing pre-PCV7 to routine-PCV7 era MLST data on carriage and IPD isolates in Native American children less than 5 years of

age (27). The results of the study indicated an increase in both carriage of non-vaccine serotypes (64% to 97% of carriage isolates) and non-vaccine serotypes isolated in cases of IPD (48% to 98% of IPD isolates) from pre-PCV7 to routine PCV7 era (27). The study also reported a significant increase in the diversity of the non-vaccine serotype 19A, a type that continues to be an important cause of IPD in the routine PCV7 era (27). A study conducted in Massachusetts among children less than 5 years of age during the routine PCV7 era gave similar results, finding 98% of pneumococcal isolates from carriage to be non-PCV7 types, with 19A the most common type, followed by 6C, and 15B/C (28).

Independent of the vaccine effect, the distribution of IPD by serotype varies over time, as reported in several studies covering the pre-vaccine era, and this natural variation must be taken into account when examining the effect of the conjugate vaccine on incidence of IPD in the population (20-31). Indeed, because of this variability in distribution of disease by serotype, Black comments that vaccine efficacy is best measured in total reduction of disease incidence (32). Nonetheless, as PCV13 is rolled out in South Africa, knowledge of pre-vaccine serotype 1 IPD epidemiology could provide a useful context in which to measure vaccine efficacy by any metric (33).

2.5 Age Specific Factors

IPD can display strong patterns across age, depending on serotype. Neonates and the elderly face the highest incidence of morbidity and mortality from invasive pneumococcal disease (1). Serotype 1 is unique in that it tends to cause infections at a regular rate throughout the first few years of life, unlike most invasive pediatric

serotypes, which predominate in causing IPD between 6 and 18 months (17). Hausdorff et al. reports that in South Africa, as in most countries, the rate of IPD due to serotype 1 remains relatively constant when moving from children younger than 4 to the 5-17 age group, whereas the rate of IPD due to the PCV7 serotypes drops dramatically (1). In adults, infection due to the PCV7 serotypes is less common, but can occur through transmission by children or when the adult is immunocompromised (1). Serotype 1 was historically responsible for large outbreaks of pneumonia that affected adults as well as children, and remains an "outbreak serotype," though these occur less frequently and only in certain circumstances (see section 2.7) (1).

2.6 HIV

The prevalence of HIV in South Africa is one of the highest in the world, with estimates of 10.5% nationwide in 2010, with KwaZulu-Natal and Gauteng provinces consistently sharing the highest disease burden (34). The serotypes contained in the PCV7 vaccine are more often isolated among HIV–infected IPD patients relative to those who are HIV-uninfected (1-2). According to Jones et al., isolation of serotype 1 from HIV-infected patients may be less likely than isolation from HIV negative patients (2). In this study, isolation of serotype 1 from bacteremic patients at a Johannesburg hospital was significantly less frequently isolated from HIV-seropositive adults, adolescents, and children, compared with HIV-seronegative individuals ($p=0.03$ for children, $p=0.009$ for adults) (2). Feiken et al. found similar results among bacteremic patients in rural Kenya where serotype 1 was the most common isolate (3).

Though there appears to be a pattern of less frequent isolation of serotype 1 from HIV positive patients, the reasons for this have not been elucidated. Feikin et al. discuss the possibility of HIV allowing for opportunistic infection by the serotypes commonly isolated in carriage, resulting in a lower attack rate for serotype 1 among HIV positive patients (3).

2.7 Outbreaks

In the pre-vaccine era, pneumonia outbreaks were commonly caused by serotype 1 (1). The transmission of pneumococcal disease occurs person to person through contact with infected droplets, which is exacerbated by poor hygiene and crowded conditions (9). Outbreaks due to serotype 1 have been reported in institutional settings (such as homeless shelters or schools), underdeveloped or rural settings, and developed communities (35-39). Notably, Mercat et al. reported in a homeless shelter outbreak that a high number of cases were alcohol users or current smokers (82% alcoholics, 90% smokers) (35). Both of these exposures are known to be risk factors for infection with *S. pneumoniae*.

It has been mentioned briefly that serotype distribution varies temporally (see section 2.4). Thus outbreaks of IPD due to serotype 1 must be considered not only as a case incidence above endemic levels, clustering according to space and time, but also as the percent contribution of the serotype to total IPD. Temporal variability can be irregular, as reported by Rückinger et al., or regular as shown by Harboe et al. (30-31). IPD also has a predictably regular seasonal variation, which additionally informs whether a true outbreak has occurred (40).

2.8 Anti-Retroviral Therapy and Cotrimoxazole prophylaxis

The South African Department of Health recommends that standard antiretroviral therapy in treatment-naïve adults and adolescents consist of triple therapy (HAART), given either when symptomatic or when CD4 count is less than or equal to 200mm^3 except if the patient is in stage IV disease (41). This threshold is higher for patients who are pregnant women or co-infected with tuberculosis at less than or equal to 350mm^3. HAART should consist of one non-nucleoside reverse transcriptase inhibitor, and two nucleoside reverse transcriptase inhibitors (41). For children, the threshold for starting antiretroviral treatment is still higher, recommended regardless of CD4 count for children less than or equal to 1 year old, and 750mm^3 for children from 2 to 4 years old (42).

HIV increases the risk of IPD at least 10-fold (43). Nunes et al. conducted a retrospective study examining the impact of Highly Active Antiretroviral Therapy (HAART) on the risk of IPD among HIV-infected children less than 18 years of age living in Soweto. The findings of the study suggest a protective effect of HAART on IPD incidence, with a parallel decrease in incidence for both PCV7 and non-PCV7 serotypes across the period of interest, moving from early to established HAART (43). Nonetheless, the overall burden of IPD did not change throughout the study period, and HIV-infected children remained at higher risk for IPD than uninfected (43).

A study conducted by Yin et al. on invasive pneumococcal disease from 2000-2009 among HIV-infected English and Welsh adults greater than 15 years of age found similar results. The average yearly incidence of IPD among subjects not receiving antiretroviral therapy was higher (281 per 100,000) than the overall average incidence (245 per 100,000) during the period of interest (44). Additionally, the proportion of among HIV-infected adults caused by serotypes 1, 5, and 7F was significantly lower than

the proportion caused by PCV7 serotypes (44). When taken together with the results from Nunes et al., this suggests that antiretroviral therapy may have a disproportionate effect on IPD caused by PCV7 serotypes, due to the positive association with IPD. The observation by Nunes et al. that a parallel decrease occurred in both PCV7 and non-PCV7 types across the period of interest, the proportion of IPD associated with PCV7 types was consistently greater than non-PCV7 among HIV-infected children, a trend not observed among HIV-uninfected children.

Cotrimoxazole prophylaxis, used to prevent HIV-associated opportunistic bacterial infections, can also decrease incidence of IPD either by itself or in combination with anti-retroviral therapy. A study conducted by Everett et al. documenting trends in IPD during antiretroviral scale-up in Malawi from 2000-2009 showed a significant decrease in IPD during that period (45). This was also observed during an era of widespread cotrimoxazole prophylaxis, but the study was unable to attribute the relative contribution of each to the decrease in IPD (45). Using a proportional hazards model, A 2010 cohort study conducted among South African adults older than 18 years of age found a significant protective effect (HR=0.64, $p<0.001$) against mortality when antiretroviral therapy was added to cotrimoxazole prophylaxis (46). Though this study, conducted by Hoffman et al., does not address disease incidence, when taken together, both Hoffman et al. and Everett et al. provide good evidence that both antiretroviral therapy and cotrimoxazole prophylaxis contribute to positive health outcomes in a high-risk environment for IPD, namely a HIV-infected populations. Therefore cotrimoxazole prophylaxis must be considered when examining the potential effect of the antiretroviral therapy on IPD incidence or mortality, and vice versa.

2.9 Relationship Between Serotype and Clinical Outcomes, including Mortality and Severity of Disease

A study by Alanee et al used a multivariable logistic regression technique to examine the association between the outcome of IPD disease and its severity, ultimately inferring risk factors for mortality (47). In particular, the study found that in multivariable analysis, underlying conditions, divided into immunosuppressive, lung, and chronic presentations, were significantly associated with mortality due to IPD, along with other factors such as age (47). The study did not find an association between IPD mortality and serotype, however in univariate analysis; increased severity seemed to be directly related to pediatric serotypes, that is, the serotypes contained in the PCV7 vaccine (47). Multivariable analysis revealed a positive association between increased severity of condition and meningitis, suppurative lung condition and underlying lung disease (47). Alanee et al. found an association in univariate analysis between nosocomial infection and mortality due to IPD, but this association was not found using a multivariable model (47).

A more recent study by Weinberger et al. using pooled estimates of risk ratio across geographic location provides evidence that mortality in pneumococcal pneumonia with bacteremia is related to serotype in contrast to the Alanee et al. study (48). In the study, the reported pooled risk of mortality due pneumococcal pneumonia caused by serotypes 1, 7F, and 8 was significantly less than serotype 14. This type was chosen as a comparison group because it was isolated in all cases of mortality across geographic location. For meningitis, no significant difference was seen in morality risk across serotype relative to type 14. The association between serotype and death was found to be unbiased by underlying co-morbidities, but strongly associated with degree of encapsulation, suggesting that properties of the serotype itself influence mortality (48).

Though factors such as underlying conditions or co-morbidities may indeed be related to the mortality of IPD as reported by Alanee et al, these findings suggest the relationship between serotype and mortality is more direct depending on clinical syndrome.

3. Intended/Potential Use of Study Findings:

This study was done with the intended purpose of partially fulfilling the requirements for the Master of Public Health degree at Emory University. The results of this study were intended to contribute the field of public health by furthering knowledge of serotype 1 invasive pneumococcal disease epidemiology.

4. Study Design/Location:

This study was conducted at the National Institute for Communicable Diseases (NICD) in Johannesburg, South Africa. The NICD is responsible for national surveillance of reportable infectious diseases in South Africa. To this end, the organization uses the Group for Enteric, Respiratory and Meningeal disease Surveillance-South Africa (GERMS-SA) network for active lab-based disease surveillance. The Respiratory and Meningeal Pathogens Research Unit (RMPRU) is primarily responsible for laboratory confirmation of *Neisseria meningitidis*, *Haemophilus influenzae,* and *S. pneumoniae* isolates from GERMS-SA sites.

Under the supervision of the RMPRU, a case-control methodology was implemented, defining serotype 1 IPD as the case, and a pool of high frequency IPD-causing serotypes as controls. This formed the basis of a "case-case" study design. The project was conducted on a de-indentified dataset based on information collected by the

unit from case report forms (CRFs) and confirmed diagnostic laboratory samples during the period 2003-2008.

5. Objectives:

1. Describe the epidemiology of serotype 1 IPD with particular respect to putative outbreaks according to time and place during the period 2003-2008 (prior to the introduction of the PCV7 vaccine in South Africa). Time will be considered as year and month of reported case. Place will be determined at the level of province.
2. Perform bivariate analyses
    a. Determine the bivariate association between relevant literature-based exposures, including HIV, and serotype 1 IPD relative to a pool of the 6 most frequent serotypes in the database using n=1000 as a cutoff.
    b. Determine the bivariate association between serotype 1 IPD and outcomes of disease severity, clinical diagnosis, and/or mortality.
3. Perform multivariable analyses
    a. Use a backwards elimination method to create a multivariable logistic regression model for refined risk factors to serotype 1 IPD using a parsimonious multivariable logistic regression model with HIV as the primary exposure.
    b. Use a stepwise regression method to create multivariable logistic regression models for serotype 1 IPD as a predictor of clinically relevant outcomes including disease severity, clinical diagnosis, and mortality

6. Hypotheses:

1. Descriptive Epidemiology
    a. Cases of serotype 1 IPD will show temporal and spatial clustering characteristic of an outbreak during the period of interest.
    b. Cases of serotype 1 will display seasonal clustering, controlling for province.
    c. Cases of serotype 1 will wax and wane relative to other serotypes across time and place.
2. Bivariate Analysis
    a. Among patients with serotype 1 IPD there will be a significantly smaller proportion that is HIV-infected relative to a pool of serotypes 14, 19A, 19F, 23F, 6A and 6B IPD.
    b. Among patients with serotype 1 IPD, there will be a significantly greater proportion that is aged 5-17 relative to the control pool
    c. Among patients with serotype 1 IPD, there will be a significantly smaller proportion that has used antibiotics in the past two months relative to the control pool.
    d. Among patients with serotype 1 IPD, there will be significantly greater proportions that are both smokers and alcohol users relative to the control pool.

e. Among patients with a severe case of IPD, defined as a Pitt bacteremia score greater than 4, there will be a significantly smaller proportion that is serotype 1 relative to patients with mild/moderate IPD.
f. Among patients with meningitis, there will be a significantly smaller proportion that is serotype 1 compared to other diagnoses.
g. Among patients with an outcome of death due to IPD, a significantly smaller proportion will be serotype 1 compared those who survive, controlling for relevant variables.
h. Among HIV-infected patients with serotype 1 IPD a significantly greater proportion will be under current ARV therapy relative to the HIV-infected control pool.

3. Multivariable models
    a. A multivariable logistic regression model with serotype 1 IPD as the outcome of interest will show age, HIV infection, antibiotic use, and lifestyle to be significant putative risk factors.
    b. A multivariable logistic regression model with serotype 1 IPD as the outcome of interest and HIV infection as the denominator will show a significantly larger proportion of serotype 1 IPD among those on current antiretroviral therapy compared to the control pool controlling for cotrimoxazole and age.
    c. A multivariable logistic regression model with mortality as the outcome of interest will show a significantly smaller proportion of mortality among

those with serotype 1 IPD compared to the control pool, controlling for underlying condition and clinical syndrome.

d. A multivariable logistic regression model with disease severity will show a significantly smaller proportion of severe disease among those with serotype 1 IPD compared to the control pool, controlling for underlying condition, clinical syndrome, and nosocomial infection.
e. A multivariable logistic regression model with meningitis as the outcome will show a significantly smaller proportion of serotype 1 IPD compared to the control pool, controlling for underlying condition and age.
f. A multivariable logistic regression model with invasive lower respiratory tract infection as the outcome will show a significantly smaller proportion of serotype 1 IPD compared to the control pool, controlling for underlying condition and age.
g. A multivariable logistic regression model including bacteremia without focus as the outcome will show a significantly smaller proportion of serotype 1 IPD compared to the control pool, controlling for underlying condition and age.

III. Procedures/Methods:

1. Study Population:

1.1 Description and source of study population

The NICD draws data on cases of reportable diseases from surveillance sites throughout the country. Thus, this study drew on samples from the entire South African population. A 2010 mid-year estimate by Statistics SA placed the national population at about 50 million people (49). The population has grown at an estimated rate of about 1% since 2001 but this rate has decreased steadily by about a tenth of a percent per year. Therefore it can be said that the study population was no larger than 50 million during the period of interest. The average life expectancy is about 53 for males, and 55 for females (49). Black Africans are the majority group according to the census population group categorization, comprising nearly 80% of the total population (49). As of 2007, 71% of the population lived in a formal dwelling, 80% used electricity for lighting, and almost 90% had access to piped water (50).

For the descriptive portion of the study, I have drawn data from all GERMS-SA sites nationwide, which include diagnostic microbiology laboratories throughout the country, both in the private and public sector, which voluntarily report cases of IPD according to the specific case definition. For the analytical portions of the study, where clinical data are necessary, I have used data from hospitals that are part of the GERMS-SA enhanced surveillance site network: at least one laboratory from each of the nine provinces.

1.2 Case Definition

For the purposes of the study, a case of serotype 1 IPD is an isolate positive for *S. pneumoniae* isolated from a normally sterile site such as blood or CSF, and serotyped by the Quellung reaction. Controls are a pool of serotypes 14, 19A, 19F, 23F, 6A, and 6B IPD cases according to positive identification and Quellung reaction.

1.3 Inclusion/exclusion criteria

All cases that fit the definition of identification of *S. pneumoniae* from a normally sterile site with a viable isolate available at the reference laboratory for serotyping were included. Cases of IPD lacking serotype data, as well as non-viable isolates were excluded from analysis

1.4 Sampling, including sample size and statistical power

2,702 cases and 9,349 controls were used for case description. Power calculations were performed based on HIV as a primary exposure, given a sample size of about 610 cases and 3,147 controls. Based on data from the RMPRU, in Gauteng province during 2003, HIV-positive patients represented 75% of IPD cases due to serotype 1 and 91% of non-serotype 1 IPD cases. Extrapolating these proportions to the study population, the sample size necessary to detect an expected 16% difference in percent exposure with 80% power (α=0.05) was calculated using the Fleiss method with continuity correction. Given the estimated ratio of 5 controls for every 1 case, there would need to be at least 54 cases and 266 controls. The estimated actual sample size would give a minimum difference in percent exposure of about 7%, detectable with 90% power (α=0.01).

1.5 Consent process

Before the enhanced clinical data was collected by interview or record review from patients at enhanced sentinel sites, patients gave informed consent to participate in the surveillance. If the patient was unable to give consent, for example, if the patient is a minor and could not understand the consent form, a relative or parent gave consent, and the child gave assent. The Case Report Form (CRF) has a box indicating whether the patient has given informed consent to the use of their personal health information.

Basic demographic data collected from national diagnostic laboratory specimens and reported to the RMPRU as part of national surveillance of reportable communicable diseases were not subject to the informed consent process.

The study was declared Human Subjects research exempt from Emory University IRB review on May 25, 2011 (IRB00049583) and exempt from the University of the Witwatersrand Human Research and Ethics Committee (Medical) review on June 24, 2011 (M110647).

1.6 Audience and stakeholder participation

This study is intended for partial fulfillment of the requirements for the Master of Public Health degree at Emory University in Atlanta, GA, USA. As such, the primary audience for this study is the general academic community. The project is under the advisement of both Dr. Anne von Gottberg (Head of Unit, RMPRU, NICD) and Dr. Keith Klugman (Professor of Global Health, Emory University). Though these parties have provided direction and guidance, I have done the analyses necessary to fulfill the objectives.

This project was funded in part by the Global Field Experience Fund, which represents a combination of endowments from the Eugene J. Gangarosa Fund, the William A. Foege Global Health Fund, and the O.C. Hubert Fellowships for International Health, as well as student fundraising.

GERMS-SA surveillance was funded in part in 2003-2006 by the United States Agency for International Development's Antimicrobial Resistance Initiative, transferred via a cooperative agreement (number U60/CCU022088) from the Centers for Disease Control and Prevention (CDC), Atlanta, Georgia. In 2005-2008, the surveillance was also supported by the CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP), Global AIDS Program (GAP) Cooperative Agreement U62/PSO022901. Additional funding for surveillance work at RMPRU is obtained from the National Institute for Communicable Diseases (NICD), division of the National Health Laboratory Service (NHLS), Johannesburg, South Africa and the Medical Research Council (MRC), South Africa.

The stakeholders of these funding sources have not influenced the conduct of this research in any way.

2. Variables:

All variables for analysis were contained within an archived Access file, which was filled using EpiInfo during 2003-2008 according to the Standard Operating Protocol (SOP). This study examined the following exposures in accordance with the literature:

Variables for Analysis

1) The reported month and year of isolation of a positive *S. pneumoniae* culture according to the case definition and exclusion/inclusion criteria. Based on laboratory data from surveillance sites.
2) Occurrence of an outbreak, defined as an incidence and/or case number of IPD at least two-fold higher than endemic levels, controlling for relevant variables. Based on laboratory data from surveillance sites.
3) The South African province where the case was reported. The provinces are as follows, in alphabetical order: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo (Northern), Mpumalanga, Northern Cape, North West, and Western Cape. Based on laboratory data from surveillance sites.
4) The patient's race, using National Census categories of Black African, Coloured, Indian or Asian, and White (49), based on enhanced data recorded on CRF.
5) The age group of the patient, excluding neonates (age less than one week), defined according to the risk groups for the disease: ≤2, 3-4, 5-17, 18-24, 25-64, and 65+ (1-4), based on enhanced data recorded on CRF.
6) The patient's gender (51), based on enhanced data recorded on CRF.
7) Clinical diagnosis, defined as discharge diagnosis according to bed letter and correlated against isolation site of specimen. Included meningitis, LRTI with bacteremia, and bacteremia without focus (1). Based on enhanced data recorded on CRF.

8) HIV coinfection at time of admission as either prior diagnosis or diagnosis at time of admission (1-3). Based on enhanced data recorded on CRF.
9) Whether the patient used any antibiotics in the prior 2 months to isolation of a positive specimen (11-13). Based on enhanced data recorded on CRF.
10) Nosocomial infection, defined as isolation of viable bacterial culture 2 or more days after admission if the patient was not referred from a previous hospital (47). If the patient was referred from another hospital, nosocomial infection was defined as a viable bacterial culture isolated 1 or more days after admission. Based on enhanced data recorded on CRF.
11) Severity of clinical syndrome. "Severe illness" was defined as Pitt bacteremia score>4 (47). Values less than or equal to 4 were grouped into mild/moderate illness, according to categories used by Alanee et al (47). Based on enhanced data recorded on CRF.
12) Underlying condition predisposing the patient to IPD infection was grouped into three general categories: protein deficiency, pulmonary condition, and cardiac conditions (47). According to the GERMS-SA instruction sheet, pulmonary conditions included asthma, chronic obstructive pulmonary disease, or cystic fibrosis. Cardiac conditions included valvular disease and heart failure. Protein malnutrtion was not further specified. Based on enhanced data recorded on CRF.
13) Lifestyle, including alcohol use and current smoker (35). Based on enhanced data recorded on CRF.

14) Disease outcome, defined as patient survival or mortality due to IPD, up to 30 days after discharge, as indicated on the CRF (47-48). If the patient was marked as refusing treatment, this was considered as survival.
15) Current treatment with any antiretroviral drugs among HIV-infected patients, considered as either current treatment, or no treatment. Past treatment was excluded (43-44). Based on enhanced data recorded on CRF.
16) Cotrimoxazole prophylaxis prior to admission among HIV-infected patients, considered as any cotrimoxazole prophylaxis, or none (45-46). Based on enhanced data recorded on CRF.

2.1 Surveillance and Laboratory Methods

The data used for this study was gathered as part of a clinical laboratory-based active surveillance network. For enhanced data, this study used information collected from CRFs and confirmed laboratory samples which are entered into a database according to SOP, as mentioned previously.

Laboratory confirmation of a *S. pneumoniae* isolate occurs in several steps. A bacterium is suspected to be *S. pneumoniae* based on characteristic alpha-hemolytic mucoid or "nail head" colony morphologies. Under the microscope, the organism is a Gram-positive diplococcus (53). The RMPRU confirms that a laboratory sample is *S. pneumoniae* through a series of procedures. The optochin susceptibility test is first performed. *S. pneumoniae* is generally susceptible to this antibiotic, and a 14mm zone of inhibition should be measured to establish susceptibility (53). This differentiates the pneumococcus from *S. viridans*, which is resistant, though *S. pneumoniae* will occasionally display resistance as well. Thus a subsequent test is for solubility in 10%

sodium deoxychoalate (bile). *S. pneumoniae* colonies are soluble in a bile solution, and a positive test will give a clear tube when compared to a turbid control (53).

All isolates of *S. pneumoniae* were serotyped using the Quellung reaction, a test in which the appearance of capsular swelling occurs when a saline suspension of bacterial capsular antigens undergo a microprecipitin reaction with the corresponding antibodies. The RMPRU uses a pooling method, in which serotype is identified through process of elimination using collections of antibodies termed pools, groups, and factors, listed in order of increasing specialization. Serotype 1 is identified through a positive test for pool P, and group A. Omniserum, which contains antibodies against all known capsular polysaccharide antigens, can be used as a confirmatory test for an otherwise unidentifiable isolate of *S. pneumoniae* (54).

In South Africa, HIV serological screening tests are generally done using the enzyme-linked immunosorbent assay (ELISA), rapid tests, or less often, the Western Blot. Tests are based on detection of anti-HIV antibodies produced in response to the virus using a standard algorithm (55).

3. Data Handling and Analysis:

3.1 Analysis

Merged data files containing cases of IPD, CRF information, and patient demographic information were imported from Access/Excel into SAS 9.2. Using this program the raw data was cleaned and patient identifiers were removed.

During the process of data cleaning, each variable was checked for logical inconsistencies. MIC values for penicillin were checked against entry into the database under susceptibility and isolation site, i.e. SPENMIC was cross-tabbed against SPEN and SPECIMEN to determine if data was entered correctly. The same process was performed for erythromycin and cotrimoxazole. Where a mismatch occurred, the original records were checked and both the original database and data for analysis were corrected. Reported age was examined against the date of birth and the calculated age in days based on the date of birth, i.e. AGE was cross-tabbed against DOB and AGEDAYS. The date of the outcome was checked against the patient's age, to ensure that if the outcome was death, the patient was not listed as older than the date of death. Likewise the date of collection was not listed as before the patient was born. Patients who were reported as having a prior HIV infection where HIVPRIOR was marked "yes" if HIVNOW was marked "no" were excluded from analysis. The clinical diagnosis as reported on the CRF was checked against the site of isolation from the specimen sent to the laboratory. All logical inconsistencies were checked against the original record. If an error was found, the data was corrected in the database.

Gender and prior antibiotic use were directly recoded from the variables in the dataset as dichotomous numeric variables. This information was based on data entered on the CRF. Province and year was based on isolate data sent from laboratory surveillance sites to the RMPRU. These were coded as numeric variables corresponding to the provinces in alphabetical order and the year in temporal order. Race, based on CRF information, was coded in the same way. HIV status was coded as a dichotomous numeric variable, either infected or uninfected. This was based on infection status at

admission as marked on the CRF, based on prior diagnosis or diagnosis at admission. Race, province, and year were also recoded dummy variables for multivariable analysis.

A continuous age variable was created, coded as the difference between birth date and date of admission as recorded on the CRF. If the continuous age was less than one week, this data was excluded from analysis. This variable was then categorized into under 2 years old, 3-4, 5-17, 18-24, 25-64, and older than 64. If the continuous age was less than one week, this data was excluded from analysis. This variable was coded as a numeric and as a series of dummy variables for multivariable analysis.

Underlying conditions included protein malnutrition, pulmonary, and cardiac conditions compared to no underlying condition, coded from information on the CRF. Any underlying condition not described was excluded. Lifestyle included smoker, alcohol user, neither, and both, coded as a numeric variable directly from information on the CRF. Similarly to age category, province, year and race, these variables were coded as a numeric variable and also as a series of dummy variables for multivariable analysis.

Nosocomial infection was considered as isolation of viable bacterial culture 2 or more days after admission if the patient was not referred from a previous hospital. If the patient was referred from another hospital, nosocomial infection was defined as a viable bacterial culture isolated 1 or more days after admission. This was based on enhanced data recorded on CRF, and coded as a dichotomous numeric variable.

Disease outcome was made dichotomous by merging discharge with refusal of treatment into "survival." CRF data indicating a severity score greater than 4 was considered severe disease, and a severity scored Clinical diagnosis was based on discharge diagnosis on the bed sheet as recorded on the CRF, and correlated against

isolation site. Lower respiratory tract infection (LRTI) was coded as such if diagnosis corresponded to isolation from lungs and blood. Meningitis was considered "meningitis" if diagnosis corresponded with isolation from CSF. Bacteremia without focus was coded as such if the isolate was taken from blood and no specific diagnosis was recorded on the bed sheet. Serotype was made dichotomous by creating a variable where serotype 1 isolates were coded the "case," and all serotypes greater than n=1000 were included as controls.

Anti-retroviral therapy was coded as a dichotomous numeric variable into current or none, based on enhanced data recorded on the CRF. Patients who had been on ARV therapy previously were excluded from analysis. Cotrimoxazole prophylaxis was also coded as a dichotomous numeric variable, as prophylaxis prior to admission or none, as recorded on the CRF.

All variables were either categorical or dichotomous, so analysis was performed using "proc freq" and "proc logistic" in SAS. A SAS macro available in the S:/ drive at the Emory University Rollins School of Public Health virtual desktop was used to assess collinearity. The number of confirmed serotype 1 IPD cases in each province per year was expressed as a raw number and as a percentage of total IPD. The raw number of serotype 1 IPD cases were then compared to the total cases of IPD in a given province from 2003-2008. Subsequently, seasonal patterns of serotype 1 IPD were described according to raw number of cases. The frequency of each serotype was calculated, and the serotypes with a cutoff of n=1000 were selected for analysis in order ensure an adequate ratio of cases to controls. Based on these criteria, a pool of the 6 most frequent serotypes after serotype 1 was used as the control group for analysis.

The estimated yearly incidence of serotype 1 IPD by province was calculated using the population of the year of analysis, as well as the reported cases of serotype 1 IPD for that year multiplied by a density factor of 1,000,000 (56). Cases reported in the previous year were subtracted from the population at risk when calculating incidence for the current year

The primary exposure of interest in the study was a co-infection with HIV at time of admission. The frequencies of serotype 1 IPD cases and non-serotype 1 controls as they are distributed across this exposure, along with the relevant covariates were analyzed for significance using Pearson chi square tests of association. The association was considered significant if the p-value was less than 0.05. The odds ratios of these variables, along with their 95% confidence intervals were calculated. If data were sparse (stratum containing <5 cases), significance and measures of association were calculated using the Mantel-Haenszel method. The distribution of disease outcome, severity, and syndrome was analyzed with serotype 1 IPD as the exposure. The penicillin, erythromycin, and cotrimoxazole susceptibility of each case was calculated, also considering serotype 1 IPD as the exposure. To help elucidate the association between reported HIV infection at time of admission and subsequent development of serotype 1 IPD, bivariate associations between cases of serotype 1 IPD and antiretroviral therapy and cotrimoxazole prophylaxis among patients who were as reported HIV infected at the time of admission was also calculated.

Once basic bivariate associations were established, a parsimonious multivariable logistic regression model was developed to establish risk factors for serotype 1 IPD. Associations between all candidate variables were calculated using chi-square tests. From

these variables the model was developed by stepwise regression using a backwards elimination approach. The stepwise regression was chosen because it combines an algorithmic approach with a strategy that allows for the discretion of the researcher (57).

Briefly, a "gold standard" model was created that regressed on the maximum number of variables possible without overloading the model or causing collinearity. Variables were removed based on their bivariate association with cases of serotype 1 IPD and other candidate variables, as well as significance of Wald t-values in the model. The point estimate and precision offered by each stepwise regression was compared to the gold standard model. The parsimonious model that offered a point estimate within 10% of the gold standard and the maximum precision was selected as the final model.

Several multivariable models were created to test the significance of serotype as a predictor of mortality, disease severity, meningitis, invasive lower respiratory tract infection, and bacteremia without focus. All initial models controlled for HIV, province, race, age, underlying condition, lifestyle, susceptibility to penicillin, erythromycin, and cotrimoxazole, antibiotic use, and nosocomial infection. The initial model with an outcome of mortality included disease severity and the three primary clinical diagnoses of interest as covariates. The initial model with an outcome of disease severity included the three primary clinical diagnoses of interest as covariates. Backwards elimination was conducted keeping serotype in the model as a predictor to find the parsimonious model within 10% of the initial model that gave optimal precision. Variables were selected for backwards elimination based on bivariate associations and Wald t-values in the initial model.

3.2 Quality Control/Assurance

The use of an SOP for data capture and entry ensures consistent quality of the information contained in the RMPRU database. Detailed instructions were provided to data clerks regarding entry of information contained on CRFs and accompanying laboratory samples. These instructions provided an algorithmic method of dealing with the complications of data entry, such as, recording of infections with multiple pathogens of interest to the RMPRU.

During the process of dataset cleaning, entries in the database were checked against original records when errors or inconsistent information appeared. Recoded variables were checked against the originals to ensure they have been created correctly. Once the final dataset for analysis was created and all errors were minimized to the greatest extent possible, the original dataset with linkers to the dataset for analysis under the custody of Dr. Anne von Gottberg. I was not and will not be able to access this information under any circumstance.

3.3 Faculty and Ethical Approval

This study was subject to the review and approval of Dr. Keith Klugman and Dr. Anne von Gottberg, the Emory University Institutional Review Board, and the University of the Witwatersrand HREC (Medical). Ethical approval was obtained from Emory University on May 25, 2011 (IRB00049583) and from the University of the Witwatersrand on June 24, 2011 (M110647).

3.4 Handling of Adverse Events:

As a retrospective database analysis, this study posed little to no risk to its subjects. In order to ensure the privacy of the subjects' personally identifiable health

information, the dataset used for analysis was removed of identifiers, with unique IDs recoded and the linking log placed in the custody of Dr. Anne von Gottberg.

IV Results:

1. Descriptive Analysis

1.1 Basic epidemiological characteristics

Serotype 1 was the most common serotype (n=2701) isolated over the six-year period (Figure 1). The seven most frequent serotypes in the dataset were 1, 14, 6A, 6B, 23F, 19A, and 19F (Figure 1). Gauteng (GA) province reported the highest frequency and rate of serotype 1 IPD cases for each year of analysis, with a peak of 43 cases per million people in 2003, and gradual decline until incidence rose to 25 cases per million people in 2008 (Figure 2). Though both Free State (FS) and KwaZulu-Natal (KZN) shared similar epidemic curves, KZN had the sharpest peak in terms of raw cases in 2005, whereas FS had the sharpest peak in terms of incidence per million people in 2004 (Figure 4). In both FS and Northern Cape (NC) provinces, the number of serotype 1 IPD cases show a marked decrease whereas total IPD shows a marked increase, particularly from 2003-2007 (Figure 6). In contrast, KZN and Western Cape (WC) show a wax and wane in serotype 1 IPD roughly parallel to total cases of IPD (Figure 6). Stratifying by serotype, 19A was the dominant serotype in WC, increasing relative to serotype 1 from 2005-2007 (Figure 7). For KZN and NC, serotypes 14 and 19F, respectively, become dominant over serotype 1 (Figure 7). In FS no one serotype exhibits this phenomenon over serotype 1, instead each serotype reaches roughly the same case number in 2008, though 14 and 23F in particular show a steady increase over the time interval (Figure 7).

1.2 Seasonality

Regular peaks in raw number of serotype 1 IPD cases was observed in GA province, occurring from August to October of each year, roughly corresponding to late

winter and early spring (Figure 8). These peaks tended to occur with less regularity in KZN, but these still largely fell within the August-October time bracket (Figure 8). Notable exceptions include a peak in July 2004, and irregular peaks throughout 2005.

2. Bivariate Analysis

2.1 Hypothesized unique risk factors for serotype 1 IPD

Significant associations were observed between cases of serotype 1 IPD and all exposures. The odds ratio of prior HIV status among serotype 1 cases relative to the control serotype pool was 0.54 (95% CI: 0.45, 0.67) (Table 1). The odds of serotype 1 IPD cases originating from Western Cape were 0.19 times the odds of the controls (95% CI: 0.15, 0.23), using Gauteng as a reference group. The odds of a case of serotype 1 IPD being both a smoker and alcohol user was about 2 times the odds for controls (95% CI: 1.51, 3.90), using a group of cases that were neither alcohol users nor smokers as a reference (Table 1). Protein deficiency showed an OR of 0.19 (95% CI: 0.11, 0.32) for cases relative to controls, using "no underlying condition" as a reference group.

2.2 Antiretroviral therapy and cotrimoxazole prophylaxis

Among HIV-infected patients with confirmed cases of IPD, 15% were on ARV therapy. Consideration of ARV therapy as an exposure gave an odds ratio of 0.14 (95% CI: 0.07, 0.28) for serotype 1 IPD relative to controls (Table 2). The proportion of HIV-infected patients with confirmed cases of IPD on cotrimoxazole prophylaxis prior to

admission to the hospital was three times that of HIV-infected patients currently on ARV therapy.

2.3 Clinical syndrome, disease severity, and mortality

Significant bivariate results were observed using serotype 1 as an exposure for clinical syndrome ($p<0.01$) and disease severity ($p=0.04$). Among patients with an outcome of mortality, the odds of confirmed serotype 1 IPD was about 1.14 times that of patients with an outcome of survival, but this association was not significant ($p=0.08$). Among patients with severe illness, the odds of confirmed serotype 1 IPD was 1.42 times (95% CI: 1.01, 2.00) the odds among patients with mild or moderate disease (Table 3).

Significant results were also observed using serotype 1 as an exposure for meningitis, LRTI, and bacteremia without focus ($p<0.01$) (Table 3). Among patients with unfocused bacteremia, the odds of confirmed serotype 1 IPD was 0.51 (95% CI: 0.36, 0.72) times the odds among patients with a diagnosis of LRTI or meningitis, whereas among patients with LRTI the odds of confirmed serotype 1 IPD were 0.72 (95% CI: 0.63, 0.80) times the odds among patients with a diagnosis of meningitis or bacteremia without focus. Among patients with meningitis, the odds of confirmed serotype 1 IPD was 1.56 (95% CI: 1.33, 1.83) times the odds among patients with a diagnosis of LRTI or bacteremia without focus.

3. Multivariable analysis

3.1 Initial Model for Serotype 1 IPD Risk Factors

The gold standard model that included the greatest number variables without overloading the model gave an adjusted OR of 0.21 (95% CI: 0.13, 0.34) for the odds of

HIV infection at time of admission for confirmed cases of serotype 1 IPD relative to controls. After eliminating the variables for underlying condition and antibiotic use due to incompatibility with the model, HIV infection, age, province, lifestyle, nosocomial infection, and race were selected for inclusion in the gold standard model. No collinearity was observed among the variables in this initial model.

3.2 Confounding Assessment and Final Model

Province, lifestyle, nosocomial infection, and gender were all selected for elimination based on association with other candidate variables and individual Wald t-statistics in the gold standard model. Province, nosocomial infection, and gender, could all be eliminated from the model without changing the point estimate for HIV by more than 10%, suggesting that they were not confounders (Table 4). The final model included HIV, age, and lifestyle as significant explanatory variables, conditioning on race. The adjusted odds ratio for HIV in this model was 0.23 (95% CI: 0.15, 0.36). The adjusted odds ratio for 5-17 year olds using ≤2 year olds as a referent was 78.73 (95% CI: 31.61, 196.08). The adjusted odds ratio for patients who were both smokers and alcohol users was 2.02 (95% CI: 1.05, 3.90)

3.3 Analysis of antiretroviral use among HIV positive patients with IPD infection

A multivariable model was developed assessing the effect of controlling for cotrimoxazole prophylaxis, lifestyle, age, and year of analysis on the association between current ARV therapy and serotype 1 IPD, with HIV infected patients as the denominator. The model controlling for all of these variables gave an adjusted odds ratio of 0.20 (95% CI: 0.06, 0.70) compared to the unadjusted estimate of 0.14 (95% CI: 0.08, 0.28),

furthermore current ARV therapy was found to be significant given the other variables in the model (Table 5).

3.4 Multivariable models containing serotype 1 IPD as a predictor of clinical outcomes

The multivariable model containing serotype 1 IPD as a predictor of mortality while conditioning on age gave an adjusted odds ratio of 0.82 (95% CI: 0.60, 1.12) while controlling for HIV infection, underlying condition other than HIV, and a clinical diagnosis of meningitis (Table 6). Serotype 1 IPD did not remain a significant predictor of mortality given the other variables in the model.

The adjusted odds ratio estimate for serotype 1 IPD as a predictor of disease severity was 1.44 (95% CI: 0.97, 2.15), also insignificant given the other variables in the model (Table 7). The model used to predict disease severity controlled for LRTI, nosocomial infection, and age.

The final multivariable model for serotype 1 IPD as a predictor of meningitis, conditioned on race and province contained HIV infection status, underlying condition other than HIV, and antibiotic use. The adjusted odds ratio for serotype 1 IPD among those with meningitis, relative to those a diagnosis of bacteremia without focus or LRTI, was 0.60 (95% CI: 0.43, 0.85) and this estimate was significant controlling for the other variables in the model ($p<0.01$) (Table 8). This suggests that the proportion of cases diagnosed as meningitis that were due to serotype 1 IPD was significantly smaller than cases diagnosed as LRTI or bacteremia without focus, controlling for the other variables in the model.

In addition to meningitis, serotype 1 IPD was also significant as a predictor of LRTI, controlling for HIV infection, underlying condition other than HIV, age, and

antibiotic use (p=0.04). Race and province proved to be confounders of the primary exposure; therefore the model was conditioned on both. The adjusted odds ratio among those with LRTI for serotype 1 IPD was 1.42 (95% CI: 1.01, 2.00) relative to those with meningitis or bacteremia without focus (Table 9). This suggests that a significantly greater proportion of LRTI infections in the dataset were due to serotype 1 IPD relative to meningitis or bacteremia without focus when controlling for the other variables in the model.

In contrast to LRTI and meningitis, bacteremia without focus was not significantly predicted by serotype 1 IPD in a model controlling for HIV infection and nosocomial infection. The adjusted odds ratio estimate for this model was 1.01 (95% CI: 0.60, 1.71) (Table 10).

V Discussion:

1. Strengths and Weaknesses

Though this study has used a case-control methodology, it is specifically a case-case design. As a study comparing one subtype of a disease to all other subtypes in the context of a national surveillance system, controls cannot necessarily represent the counterfactual experience of cases (6). Though this methodology restricts analysis from examining healthy individuals, the use of a multivariable logistic regression model allows for estimating refined risk factors for exposures specific to the outcome of interest, that is, serotype 1 IPD relative to the control pool (6). Furthermore, the use of a multivariable logistic regression model can provide a strong estimate the relative risk for predictors of interest, controlling for other factors (57).

The major limitation of this study has been touched on in the previous paragraph, namely that a national surveillance system cannot fully report the burden of disease in a population, nor can the study directly infer risk factors for healthy individuals. GERMS-SA enhanced sites may likewise not necessarily be representative of the whole South African population. Furthermore, the use of enhanced sites for clinical data further restricts the sample size available for analysis. Regardless of the size of the sample, only isolates that could be serotyped were analyzed.

This study faced some selection bias. Only patients who develop disease, go to a hospital for treatment, and have specimens taken for diagnosis are eligible for analysis by the RMPRU. Furthermore, the ability of the RMPRU to analyze such specimens is dependent entirely on the diagnostic laboratory providing the specimen to the unit. A study by Crowther et al. reported that factors such as organism, province, site of

specimen collection, and age group all were significant predictors of whether a case was reported (58).

Given the evidence in the literature that serotype 1 is less likely to cause severe disease than other types, one might expect that cases of serotype 1 IPD are less often hospitalized and therefore this study has underestimated the true burden of disease due to serotype 1 IPD in South Africa. Though this may be true in a sense, the identification of serotype is unbiased because isolates are only typed after all associated clinically relevant variables are collected. It is possible, nonetheless that results of multivariable analysis suggesting that for this dataset serotype 1 was not a significant predictor of disease severity, may be a result of this bias.

A significant source of bias came in potential misclassification of categorical exposures. To address this, the data were grouped as closely as possible to categories found in the literature. Potential confounding of the association between the outcomes and the exposures was addressed in the design of the study.

2. Discussion of Findings

These results attempt to elucidate the epidemiology of serotype 1 IPD through descriptive, exploratory and confirmatory analyses. The distribution of serotypes in the data indicates that the seven most common IPD-causing serotypes during the period of interest are all contained in PCV13. This has positive implication for coverage of the vaccine. The case data suggests the same temporal fluctuation in serotype discussed by in multiple previous studies (29-31). In many provinces, dips in case number of serotype 1 IPD corresponded to a rise in another serotype. In this dataset serotype 14 was commonly but not exclusively such an example. Observational evidence suggests serotype 1

outbreaks during 2003-2008 followed a distinct seasonal pattern, however this was not strictly observed across all provinces for each year. This pattern was most clearly observed in Gauteng province. This makes sense given a number of factors, including larger population, increased number of surveillance sites, and smaller total area relative to other provinces. Taken together, these factors contribute to increased resolution of data in Gauteng province. With this in mind, departures from a seasonal pattern could be a result of differences in climate (e.g. comparing KwaZulu-Natal to Gauteng to Western Cape), sparse data (e.g. Limpopo, Northern Cape), or unseasonable outbreaks due to other factors. Time series analysis may be useful in shedding light on the statistical significance of this data.

The observation that a significantly smaller proportion of HIV-infected patients with a diagnosis of serotype 1 IPD were undergoing antiretroviral treatment relative to the control pool merits consideration. Despite the dramatically inverse relationship (OR=0.14) observed in the unadjusted association between antiretroviral therapy and serotype 1 IPD, several factors lend caution to the conclusion that ARV has a protective effect on IPD due to serotype 1. Primarily, the effect of controlling for cotrimoxazole prophylaxis, lifestyle, and age suggests the unadjusted association is confounded by these factors, biasing the estimate away from the null. The addition of more covariates may eliminate this association. Furthermore the proportion of HIV positive patients with a diagnosis of serotype 1 is significantly smaller than the in the control pool, and the proportion of those patients with access to ARV therapy is smaller still. Therefore the association observed would not represent a causal link at all, but rather indicates a specific characteristic of both the denominator and the exposure, specifically that

isolation of serotype 1 during IPD is rare enough among the HIV-infected, and the proportion of those on ARV therapy is small enough to result in a statistically significant association. Finally, though ARV therapy was used throughout the period of interest, prevalence of use increased from 2003-2008. This analysis attempts to take this temporal trend into account by conditioning the logistic regression model on year, and the results suggest that year is indeed a confounder of the association between serotype 1 IPD and current ARV therapy.

The association of HIV infection at time of admission with serotype 1 IPD is in accordance with the findings of Jones et al. and Feiken et al. (2-3). Similarly the association of age with serotype 1 IPD generally agrees with the analyses of Hausdorff et al., particularly the finding that children aged 5-17 have a significantly higher putative risk for serotype 1 IPD relative to the control pool, even when controlling for relevant variables (1).

The direct positive association between serotype 1 IPD and exposure to both alcohol and smoking is noteworthy given that these are risk factors for IPD regardless of serotype. Though the results of this study suggest smoking/alcohol use is a refined risk factor for serotype 1 IPD, this does not preclude it from being a general risk factor for IPD, it only suggests that the IPD is likely to be due to serotype 1. This is supported by studies finding alcohol use and smoking to be characteristic of cases in pneumococcal outbreaks, given that serotype 1 is often isolated in outbreak situations. It is also possible that these behaviors may have a synergistic effect on infection with a serotype exhibiting a high invasiveness index, such as serotype 1. This is supported by the fact that when controlling for race, age, and HIV infection, the putative risk of infection with serotype 1

IPD among patients who used both alcohol and smoked remains two times higher than patients who did not report this behavior, and this risk is higher than patients who either smoked or used alcohol, but not both. Further studies could improve the resolution of these results by specifying frequency of alcohol use and smoking behaviors.

It is important to note that increased susceptibility to antibiotics is affected by less antibiotic use, which is also significant in the results of the bivariate and multivariable analysis. The finding of this study that there is significantly less antibiotic use among those with serotype 1 IPD would supported by the corresponding finding that serotype 1 is more susceptible to antibiotics. Alternatively, given that serotype 1 is more susceptible, it would make sense that antibiotic use would result in less frequent isolation from blood or other sterile sites. Further studies could assess the antibiotic susceptibility of this data.

The observation in bivariate analysis that a significantly smaller proportion of serotype 1 IPD isolates were from nosocomial infections compared to the control pool could be confounded by HIV, given the high proportion of patients with nosocomial IPD that were infected with HIV at the time of admission (70%). A significant association was observed between HIV and nosocomial infection in the confounding assessment portion of analysis, which lends further support to this hypothesis.

The contrast between results of bivariate and multivariable analysis in respect to clinical outcomes of interest suggests that several factors confound the bivariate association. These variables generally were HIV, underlying condition other than HIV, antibiotic use, and age category. Despite the fact that a multivariable model seemed to confirm the relationship between serotype 1 IPD and mortality hypothesized by

Weinberger et al., this relationship was not significant, suggesting that underlying condition, and clinical diagnosis was more important in predicting clinical outcome for this dataset, given the way the variables were defined. This finding would seem to support the findings of Alanee et al.

Even when controlling for age, nosocomial infection, and diagnosis of LRTI, the odds of isolating serotype 1 in cases of severe IPD was about equal to the odds of isolation in mild or moderate cases of IPD, and this relationship was insignificant. Even when controlling for these variables, the adjusted estimate is about equal to the unadjusted estimate, however in contrast to the unadjusted estimate, the adjusted estimate is insignificant. This suggests that nosocomial infection, LRTI, and age are better predictors of disease severity in the dataset, but are not confounders of the relationship between serotype and disease severity. Therefore there may be a confounder that was not controlled for in the model that influences the observed relationship in the dataset.

The results of this analysis suggest that unlike serotype 1, HIV infection along with other underlying conditions are better predictors of bacteremia without focus relative to meningitis and LRTI. These factors also apparently confound the bivariate relationship between bacteremia without focus and serotype 1 IPD. The contrast between the bivariate and multivariable association between serotype 1 IPD and meningitis and LRTI respectively suggests that these relationships are significantly confounded by HIV-infection, underlying condition other than HIV, antibiotic use, age, race, and province. Furthermore the fact that the respective associations between serotype 1 IPD and a clinical diagnosis of LRTI/meningitis remain significant after controlling for these variables suggests that serotype 1 IPD is significantly more likely to be LRTI relative to

meningitis or bacteremia without focus and less likely to be meningitis relative to LRTI or bacteremia without focus for this dataset.

VI Conclusion:

Serotype 1 incidence and case number in South Africa exhibited the same waxing and waning during the period 2003-2008 observed in other studies. The proportion of total IPD exhibited by serotype 1 varied by province and was generally either parallel to or inversely proportionate to a rise and fall in total IPD. Serotype composition varied by province, and in some cases, a fall in serotype 1 corresponded to a rise in another serotype, most frequently 23F or 14. For Gauteng province in particular, 2003 and 2008 stood out as "outbreak" years, where both case number and incidence spiked relative to other years. Additionally, Gauteng province exhibited the strongest seasonal pattern of serotype 1 IPD across all provinces, though seasonal patterns were evident for most provinces. In bivariate analysis, serotype 1 was significant with all exposures of interest, but only age, HIV, race, and lifestyle were included in a multivariable model with serotype as the outcome. According to this model, refined risk factors for serotype 1 IPD include being in the 5-17 year age group, and use of alcohol in conjunction with smoking. Among HIV-infected patients, the proportion of patients on antiretroviral therapy was significantly smaller among those with serotype 1 IPD relative to the control pool, but this was biased by cotrimoxazole prophylaxis, age, and lifestyle. Multivariable models including serotype as a primary predictor showed serotype to be an insignificant predictor for mortality, disease severity and bacteremia in this dataset. Serotype 1 IPD was a significant primary predictor of meningitis and LRTI respectively, controlling for HIV infection, underlying condition other than HIV, age, race, province, and antibiotic use.

VII References:

1. Hausdorff WP, Feikin DR, Klugman KP. "Epidemiological differences among pneumococcal serotypes." *Lancet Infectious Diseases*. 2005;5:83-93
2. Jones N, Huebner R, Khoosal M, et al. "The impact of HIV on *Streptococcus pneumoniae* bacteraemia in a South African population." *AIDS.* 1998;12(16):2177-2184
3. Feikin DR, Jagero G, Aura B, et al. "High rate of pneumococcal bacteremia in a prospective cohort of older children and adults in an area of high HIV prevalence in rural western Kenya." *BMC Infectious Diseases*. 2010;10:e186
4. Johnson HL, Deloria-Knoll M, Levine OS, et al. "Systematic Evaluation of Serotypes Causing Pneumococcal Disease among Children Under Five: The Pneumococcal Global Serotype Project." *PLoS Medicine*. 2010;7(10):e1000348
5. Brueggemann AB, Spratt BG. "Temporal and Geographic Stability of the Serogroup-Specific Invasive Disease Potential of *Streptococcus pneumoniae* in Children." *The Journal of Infectious Diseases*. 2004;190:1203-1211
6. Brueggemann AB, Griffiths DT, Meats E, et al. "Clonal Relationships between Invasive and Carriage *Streptococcus pneumoniae* and Serotype- and Clone-Specific Differences in Invasive Disease Potential." *The Journal of Infectious Diseases*. 2003;187:1424-1432
7. National Institute for Communicable Diseases. "Provisional listing of laboratory-confirmed cases of diseases under surveillance: 01 January-30 June 2010."

Communicable Diseases Surveillance Bulletin. 2010;8(3): 49. (http://www.nicd.ac.za). (Accessed June 6, 2011).

8. McCarthy N, Giesecke J. "Case-case comparisons to study causation of common infectious diseases" *International Journal of Epidemiology*. 1999;28: 764-768
9. Nelson KE, Williams CM. *Infectious Disease Epidemiology: Theory and Practice – Second Edition*. Sudbury, MA: Jones and Bartlett; 2007
10. Kadioglu A, Welser JN, Paton JC, et al. "The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease." *Nature Reviews Microbiology*. 2008;6:288-301
11. Mandell GL, Bennet JE, Dolin R. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases- Seventh Edition*. Philadelphia, PA: Churchill Livingstone Elsevier; 2010
12. Bédos JP, Chevret S, Chastang C, et al. (1996) "Epidemiological Features of and Risk Factors for Infection by *Streptococcus pneumoniae* Strains with Diminished Susceptibility to Penicillin: Findings of a French Survey." *Clinical Infectious Diseases*. 1996;22:63-72.
13. Clavo-Sánchez AJ, Gíron-González JA, López-Prieto D, et al. "Multivariate Analysis of Risk Factors for Infection Due to Penicillin Resistant and Multidrug-Resistant *Streptococcus pneumoniae*: A Multicenter Study." *Clinical Infectious Diseases*. 1997;24:1052-1059
14. Van Eldere J, Mera RM, Miller LA, et al. "Risk Factors for Development of Multiple-Class Resistance to *Streptococcus pnuemoniae* Strains in Belgium over a

10-Year Period: Antimicrobial Consumption, Population Density, and Geographic Location." *Antimicrobial Agents and Chemotherapy*. 2007;51(10): 3491-97

15. Peterson LR. "Penicillins for Treatment of Pneumococcal Pneumonia: Does *In Vitro* Resistance Really Matter?" *Clinical Infectious Diseases*. 2006;42:224-33.
16. Fenoll A, Granzio JJ, Aguilar M, et al. "Temporal Trends of Invasive Streptococcus pneumoniae Serotypes and Antimicrobial Resistance Patterns in Spain from 1979-2007." *Journal of Clinical Microbiology*. 2009;47(4):(doi: 10.1128/JCM.01454-08.).
17. World Health Organization. "Pneumococcal conjugate vaccine for childhood immunization – WHO position paper- Vaccin antipneumococcique conjugué pour la vaccination infantile – note d'information de l'OMS." *Weekly epidemiological record- Relevé épidémiologique hebdomadaire*. 2007;82(12): 93-104. (http://www.who.int/en/). (Accessed June 6, 2011).
18. Health Department- Republic of South Africa. "Expanded Programme on Immunisation – EPI (SA): Revised Childhood Immunisation Schedule from April 2009." 2009. (http://www.doh.gov.za/.) Accessed June 6, 2011.
19. Cutts FT, Zaman SMA, Enwere G, et al. (2005) "Efficacy of nine-valent pneumococcal vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomized, double-blind, placebo-controlled trial." *Lancet Infectious Diseases.* 2005;365: 1139-46
20. Klugman KP, Madhi SA, Huebner RE, et al. "A Trial of a 9-Valent Pneumococcal Conjugate Vaccine in Children With and Those Without HIV Infection." *The New England Journal of Medicine*. 2005;349(14): 1341-8

21. Klugman KP, Madhi SA, Adegbola RA, et al.“Timing of serotype 1 pneumococcal disease suggests the need for evaluation of a booster dose” *Vaccine*. 2011;29: 3372-73

22. Davis, S. “Prevention and Treatment of Pneumococcal Disease.” *Professional Nursing Today*. 2011;15(3): 38-41

23. Nunes S, Sá-Leão R, Pereira LC, et al. “Emergence of a serotype 1 *Streptococcus pneumoniae* lineage colonizing healthy children in Portugal in the seven-valent conjugate vaccination era.” *Clinical Microbiology and Infectious Diseases*. 2008;14: 82-84

24. Klugman KP. “Efficacy of pneumococcal conjugate vaccines and their effect on carriage and antimicrobial resistance.” *Lancet Infectious Diseases*. 2001;1:85-91

25. Mbelle N, Huebner RE, Wasas AD, et al. “Immunogenecity and Impact on Nasopharyngeal Carriage of a Nonavalent Pneumococcal Conjugate Vaccine.” *The Journal of Infectious Disease*. 1999;180:1171-1176

26. Weinberger DM, Malley R, Lipsitch M. “Serotype replacement in disease after pneumococcal vaccination.” *Lancet Infectious Diseases*. 2011; (doi:10.1016/S0140- 6736(10)62225-8).

27. Scott JR, Hanage WP, Lipsitch M. “Pneumococcal sequence type replacement among American Indian children: A comparison of pre- and routine-PCV7 eras.” *Vaccine*. 2011;30(2012):2376-2381. (doi:10.1016/j.vaccine.2011.11.004).

28. Wroe PC, Lee GM, Finkelstein JA. “Pneumococcal carriage and antibiotic resistance in young children before 13- valent conjugate vaccine.” [abstract] *Pediatric Infectious Disease Journal*. 2012;31(3):249-254

29. Jeffries JM, Smith AJ, Edwards FS, et al. "Temporal Analysis of Invasive Pneumococcal Clones from Scotland Illustrates Fluctuations in Diversity of Serotype and Genotype in the Absence of Pneumococcal Conjugate Vaccine." *Journal of Clinical Microbiology*. 2010;48(1): 87-96.
30. Rückinger S, von Kries R, Reinert RR, et al. "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;26:3984-86.
31. Harboe ZB, Benfield TL, Valentiner-Branth P, et al. "Temporal Trends in Invasive Pneumococcal Disease and Pneumococcal Serotypes over 7 Decades." *Clinical Infectious Diseases*. 2010;50: 329-37
32. Black S (2010) "The Volatile Nature of Pneumococcal Serotype Epidemiology: The Potential for Misinterpretation." *The Pediatric Infectious Disease Journal*. 2010;29(3):(doi: 10.1097/INF.0b013e3181c391fb).
33. Von Gottberg A. "Bacterial meningitis in the era of paediatric vaccination against the encapsulated pathogens." *Continuing Medical Education*. 2010;28(6):252-259
34. Health Systems Trust. "Health Statistics: HIV Prevalence (%) (total population)" 2011; (http://www.hst.org.za/healthstats/84/data). (Accessed April 15, 2012).
35. Mercat A, Nguyen J, Dautzenburg B. "An Outbreak of pneumococcal pneumonia in two men's shelters." *CHEST*. 1999;99(1):147-151
36. Gupta A, Khaw FM, Stokle EK, et al. "Outbreak of *Streptococcus pneumoniae* serotype 1 pneumonia in a United Kingdom school." *BMJ*. 2008;337: (doi:10.1136/bmj.a2964).

37. Leimkugel J, Forgor AA, Gagneux S, et al. “An Outbreak of Serotype 1 *Streptococcus pneumoniae* Meningitis in Northern Ghana with Features That Are Characteristic of *Neisseria meningitides* Meningitis Epidemics.” *The Journal of Infectious Diseases*. 2005;192(14): 192-199

38. Dagan R, Gradstein S, Belmaker I, et al. “An Outbreak of Streptococcus pneumoniae Serotype 1 in a Closed Community in Southern Israel.” *Clinical Infectious Diseases*. 2000;30:319-21

39. Le Hello S, Watson M, Levy M, et al. “Invasive serotype 1 *Streptococcus pneumoniae* outbreaks in the South Pacific (2000-2007).” *Journal of Clinical Microbiology*. 2010; (doi:10.1128/JCM.01615-09).

40. Von Gottberg A, de Gouveia L, du Plessis M, et al. “Outbreak of serotype 1 invasive pneumococcal disease in South Africa in 2003.” [abstract] *NICD Academic Day*. 2006.

41. Health Department- Republic of South Africa. “Recommended ART Regimens for Treatment-Naïve Adults and Adolescents.” *Clinical Guidelines for the Management of HIV & AIDS in Adults and Adolescents*. 2010;14-16. (www.fidssa.co.za/Guidelines/2010_Adult_ART_Guidelines.pdf). (Accessed April 11, 2012).

42. Health Department- Republic of South Africa. “Section 6: Anti-Retroviral Therapy (ART).” *Guidelines for the Management of HIV in Children: 2nd Edition*. 2010;28-35

43. Nunes MC, von Gottberg A, de Gouveia L, et al. “The impact of antiretroviral treatment on the burden of invasive pneumococcal disease in South African children: a time series analysis.” *AIDS*. 2011;25:453-462.

44. Yin Z, Rice BD, Waight, P, et al. “Invasive pneumococcal disease among HIV-positive individuals, 2000-2009.” *AIDS*. 2012;26:87-94

45. Everett DB, Mukaka M, Denis B, et al. “Ten Years of Surveillance for Invasive *Streptococcus pneumoniae* during the Era of Antiretroviral Scale-Up and Cotrimoxazole Prophylaxis in Malawi.” PLoS One. 2011;6(3):e17765

46. Hoffman CJ, Fielding KL, Charalambous S, et al. “Reducing mortality with cotrimoxazole preventive therapy at initiation of antiretroviral therapy in South Africa.” *AIDS*. 2010;24(11):(doi:10.1097/QAD.0b013e32833ac6bc).

47. Alanee SRJ, McGee L, Jackson D, et al. “Association of Serotypes of *Streptococcus pneumoniae* with Disease Severity and Outcome in Adults: An International Study” *Clinical Infectious Diseases*. 2007;45(1): 46-51

48. Weinburger DM, Harboe ZB, Sanders EAM, et al. “Association of Serotype with Risk of Death Due to Pneumococcal Pneumonia: A Meta-Analysis.” *Clinical Infectious Disease*. 2010;51(6): 692-99.

49. Statistics South Africa “Mid-year population estimates 2010.” Release P0302. 2010; (http://www.statssa.gov.za/). (Accessed June 6, 2011).

50. Statistics South Africa “Community survey 2007 (Revised Version).” Release P0301. 2007; (http://www.statssa.gov.za/). (Accessed June 6, 2011).

51. Buie KA, Klugman KP, von Gottberg A, et al. “Gender as a Risk Factor for Both Antibiotic Resistance and Infection with Pediatric Serogroups/Serotypes, in HIV-

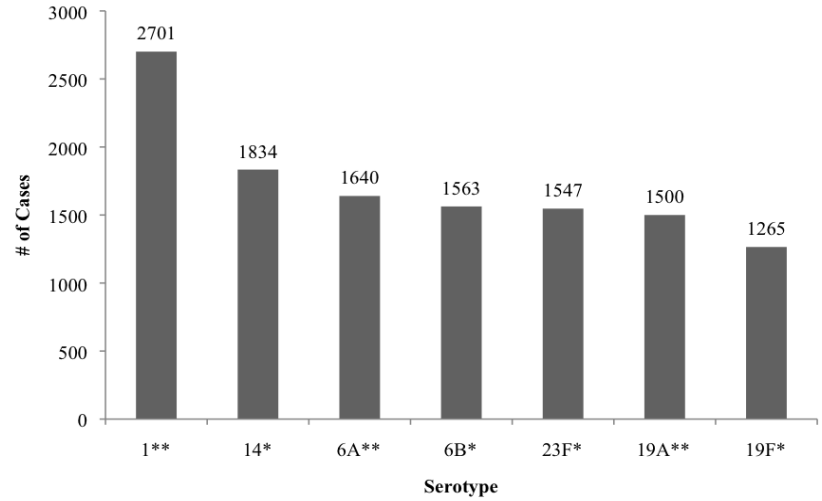
Infected and –Uninfected Adults with Pneumococcal Bacteremia." *The Journal of Infectious Diseases*. 2004;189: 1996-2000

52. CLSI "Performance Standards for Antimicrobial Susceptibility Testing; Nineteenth Informational Supplement." 2009;29(3): 72-78

53. Koneman EW, Winn Jr W, Allen S, et al. *Diagnostic Microbiology- Fourth Edition*. Philadelphia, PA: J.B. Lippincott Company; 1992.
54. Richter SS, Heilmann KP, Dohrn CL, et al. (2008) "Accuracy of Phenotypic Methods for Identification of *Streptococcus pneumoniae* Isolates Included in Surveillance Programs." *Journal of Clinical Microbiology*. 46(7): 2184-88
55. Singh E, Cohen C, Govender N, et al. "A Description of HIV Testing Strategies at 21 Laboratories in South Africa." *Communicable Diseases Surveillance Bulletin*. 2008;6(4): 16-17 (http://www.nicd.ac.za). (Accessed June 6, 2011).
56. Hajjeh RA, Sofair AN, Harrison LH, et al. "Incidence of Bloodstream Infections Due to Candida Species and In Vitro Susceptibilities of Isolates Collected from 1998-2000 in a Population-Based Active Surveillance Program." *Journal of Clinical Microbiology*. 2004;42(4):1519-1527.
57. Kleinbaum DG, Kupper LL, Nizam A, et al. *Applied Regression Analysis and Other Multivariable Methods*. Belmont, CA:Thomson Higher Education; 2008
58. Crowther P, Cohen C, Govender N, et al. (2008) "Predictors of Non-Reporting to a National Laboratory-Based Surveillance Programme." *Communicable Diseases Surveillance Bulletin*. 6(4): 1-4. (http://www.nicd.ac.za). (Accessed June 6, 2011).

VIII. Figures and Tables

Figure 1
Total reported cases of IPD by serotype



*PCV7 serotype
**PCV13 serotype

Figure 2
Reported serotype 1 (ST1) IPD cases per year in South Africa: 2003-2008

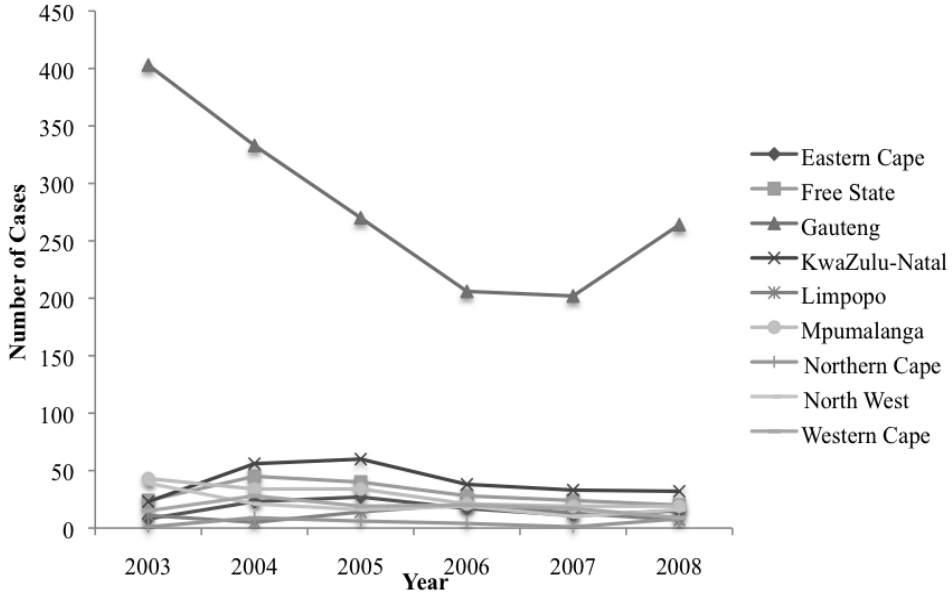


Figure 3
Estimated incidence of ST1 IPD cases per 1,000,000 people in SA: 2003-2008

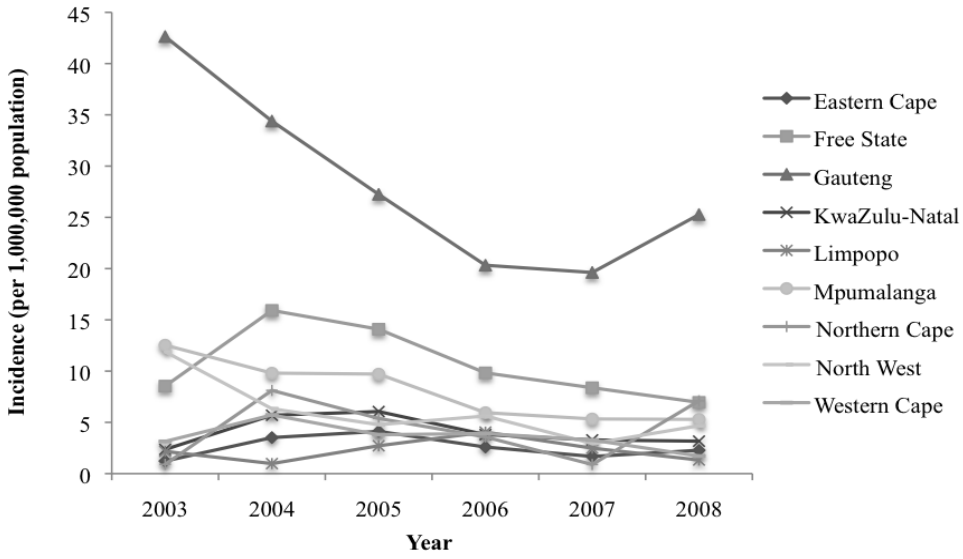


Figure 4
Reported ST1 Cases per year in SA excluding Gauteng Province: 2003-2008

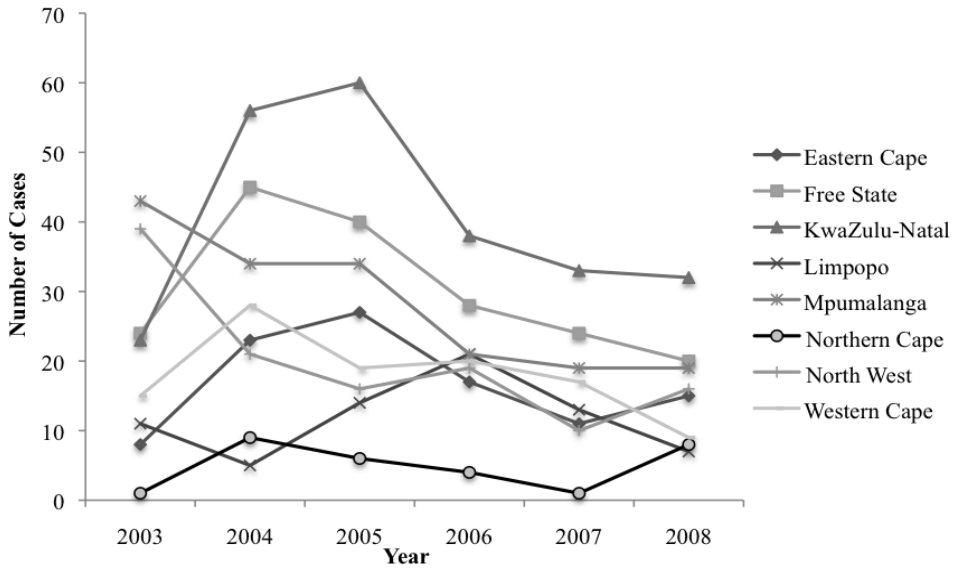


Figure 5
Estimated incidence of ST1 per 1,000,000 people in SA excluding Gauteng Province 2003-2008

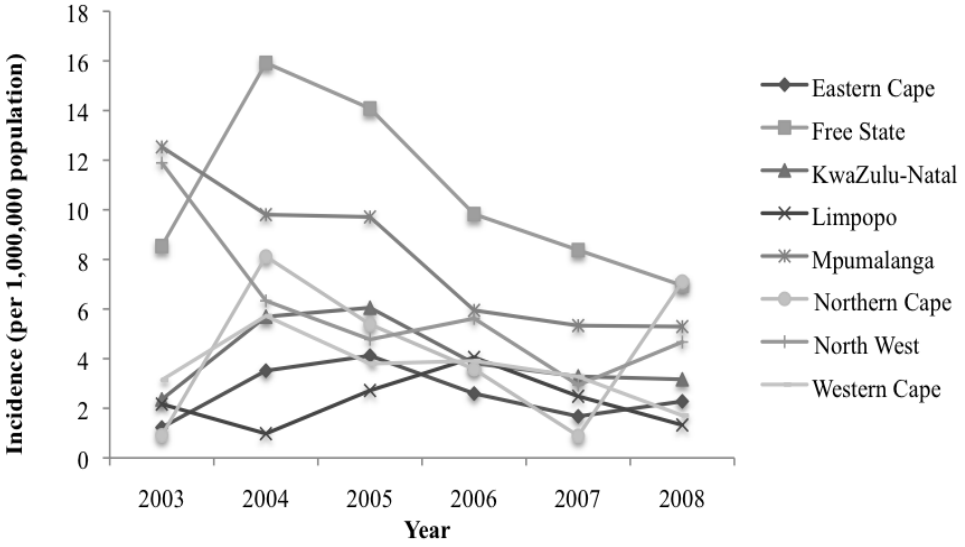
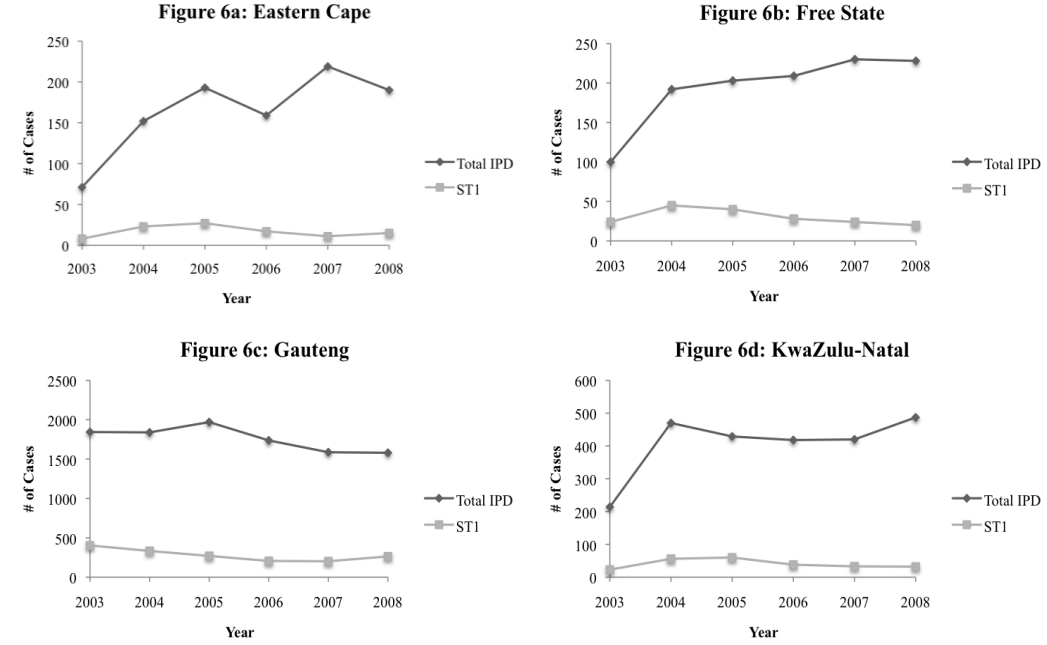


Figure 6
Reported cases of serotype 1 IPD over time relative to total IPD by province 2003-2008



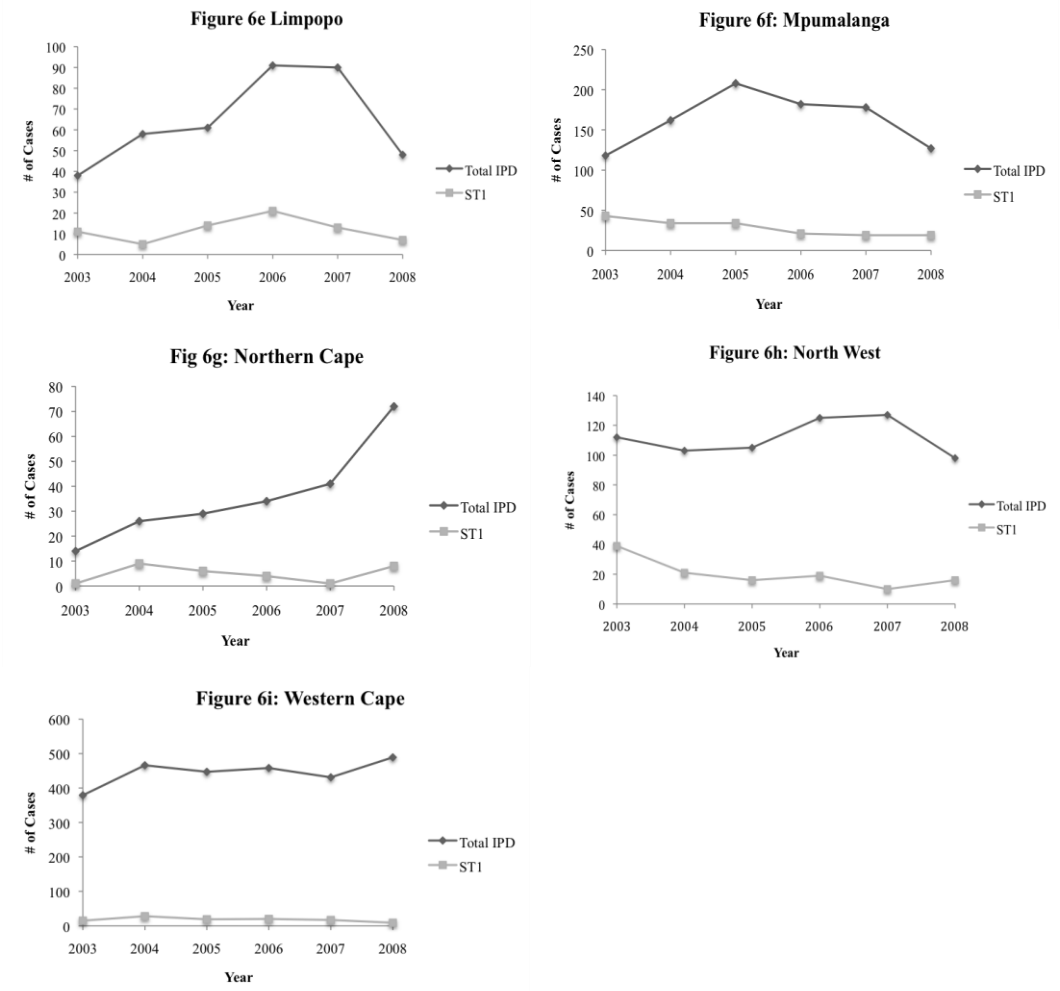


Figure 7
Reported cases of selected serotypes, by province, 2003-2008

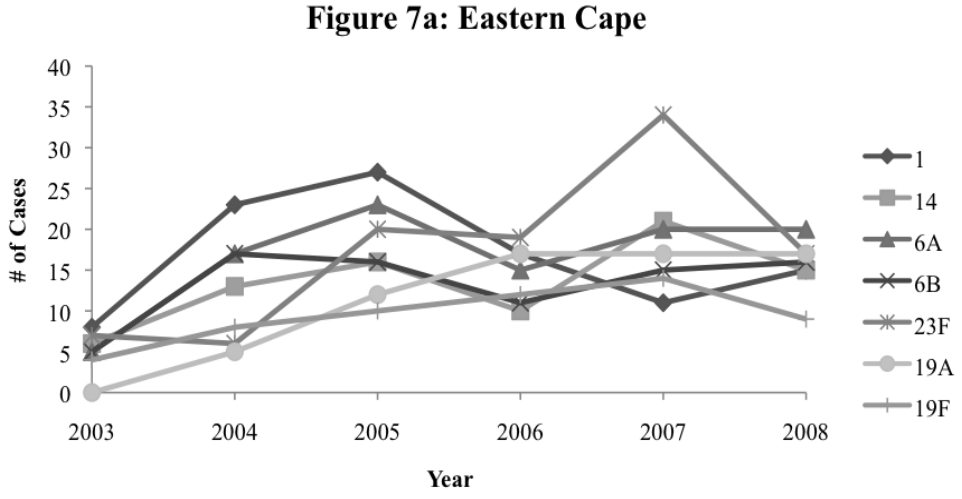


Figure 7b: Free State

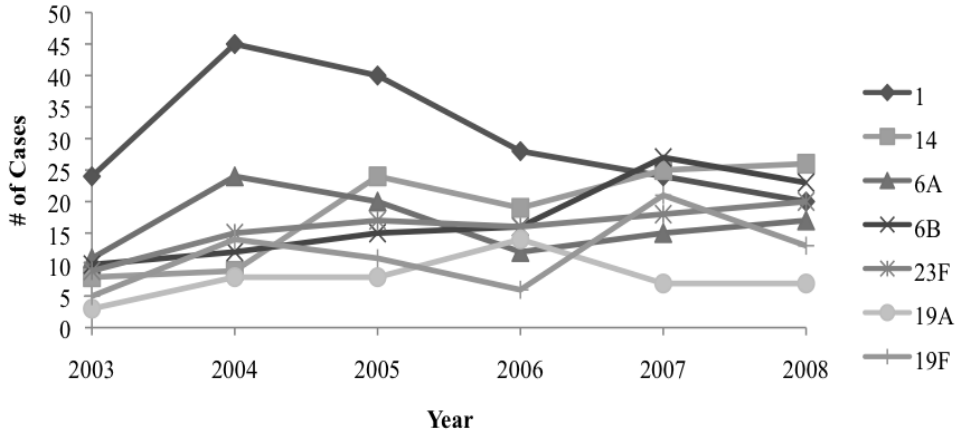


Figure 7c: KwaZulu-Natal

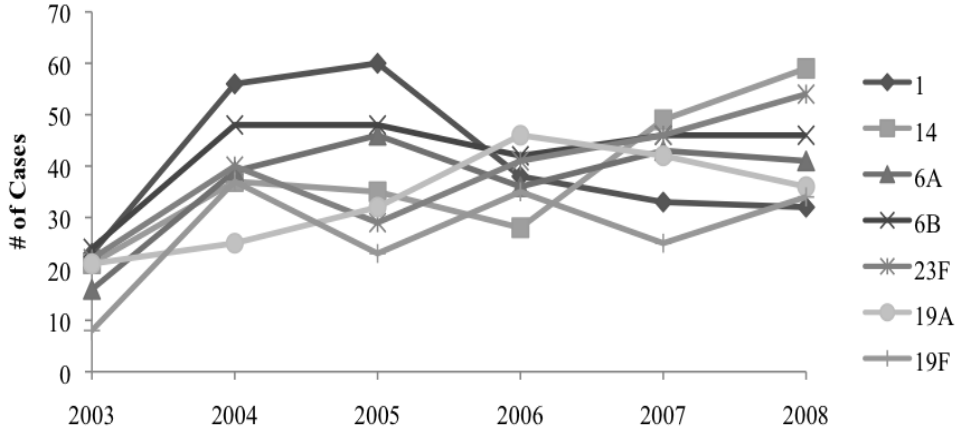


Figure 7d: Limpopo

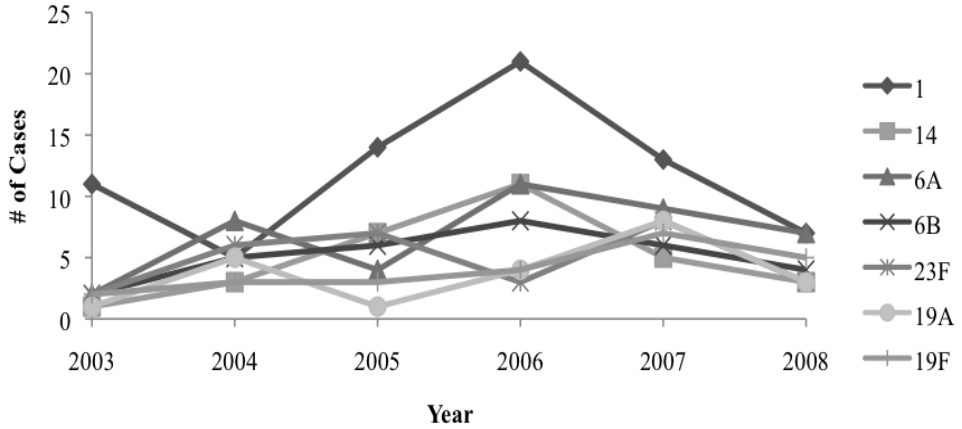


Figure 7e: Mpumalanga

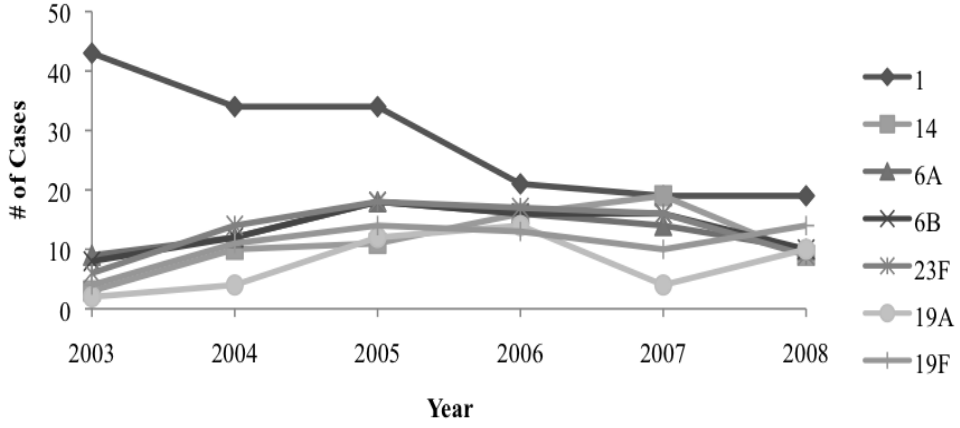


Figure 7f: Northern Cape

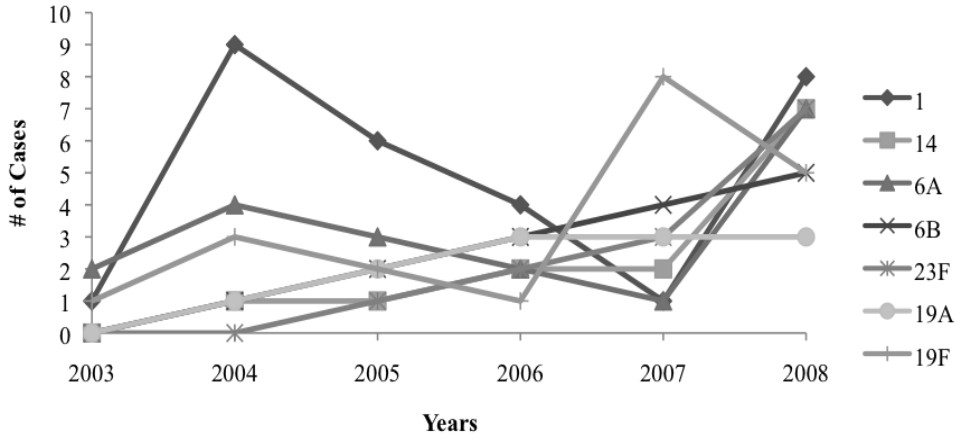


Figure 7g: North West

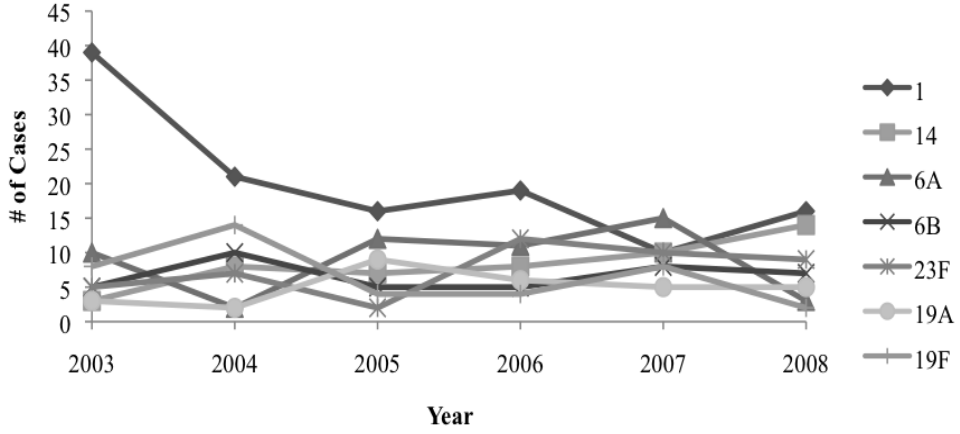


Figure 7h: Western Cape

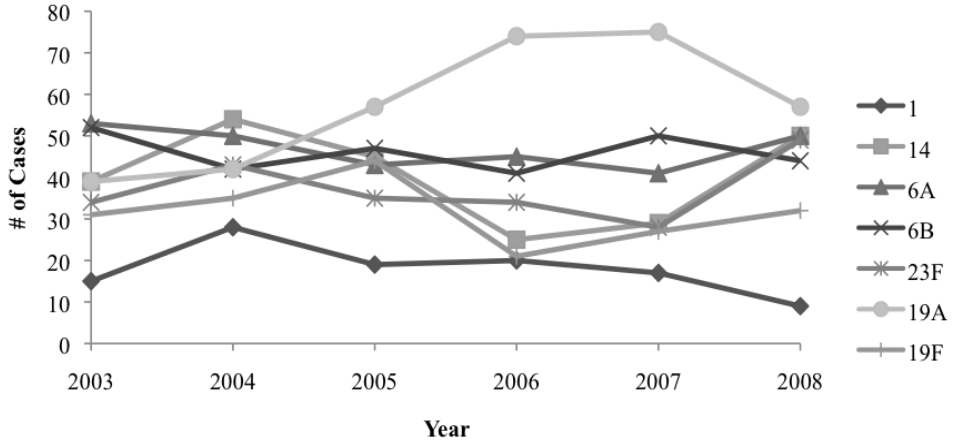


Fig 7i Gauteng

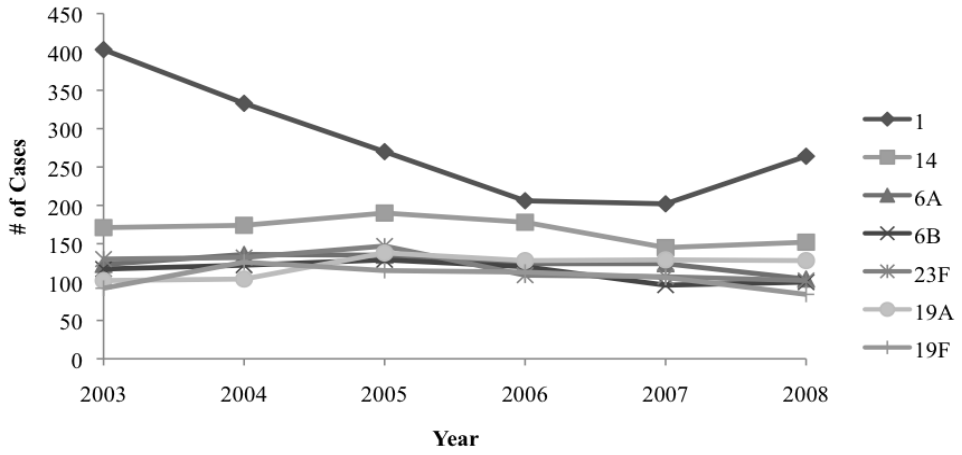


Figure 8
Monthly reported cases of serotype 1 IPD, by province, 2003-2008

Figure 8a: Eastern Cape

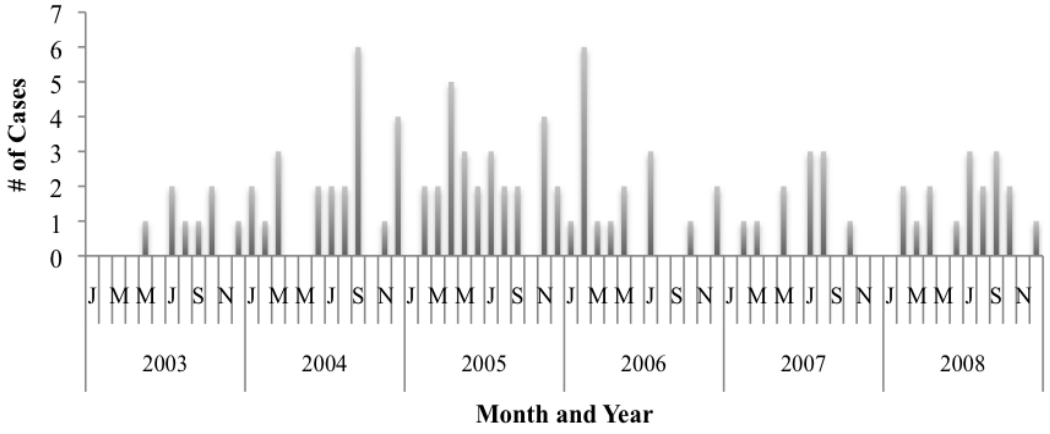


Figure 8b: Free State

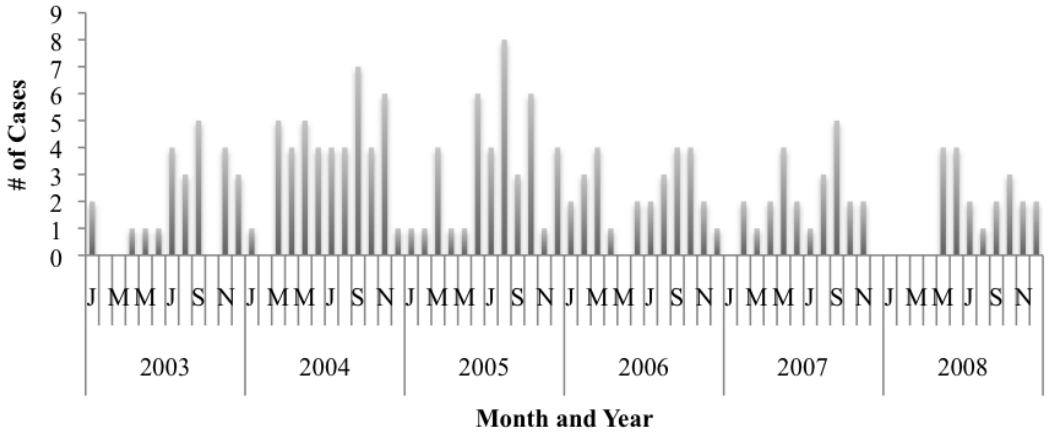


Figure 8c: Gauteng

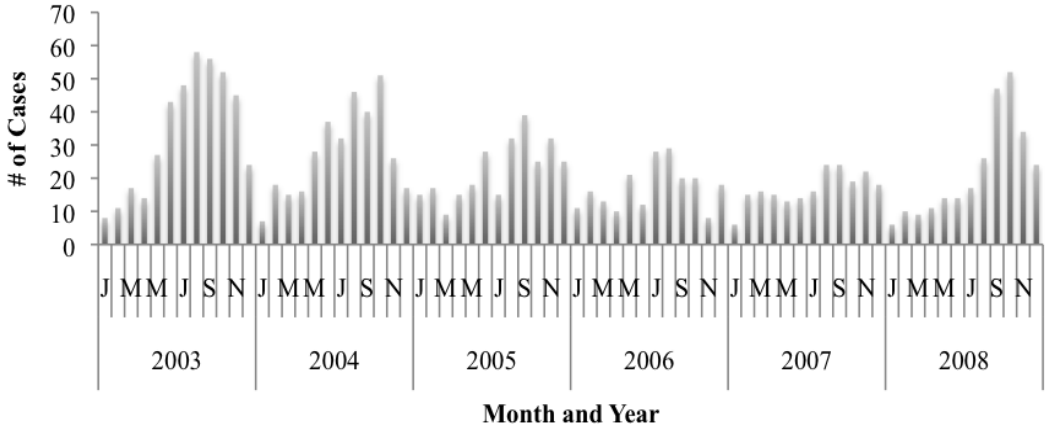


Figure 8d: KwaZulu-Natal

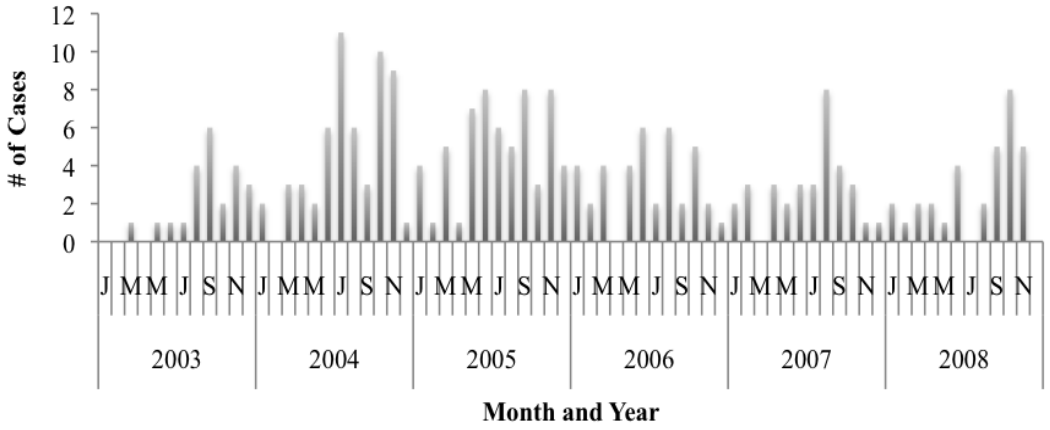


Figure 8e: Limpopo

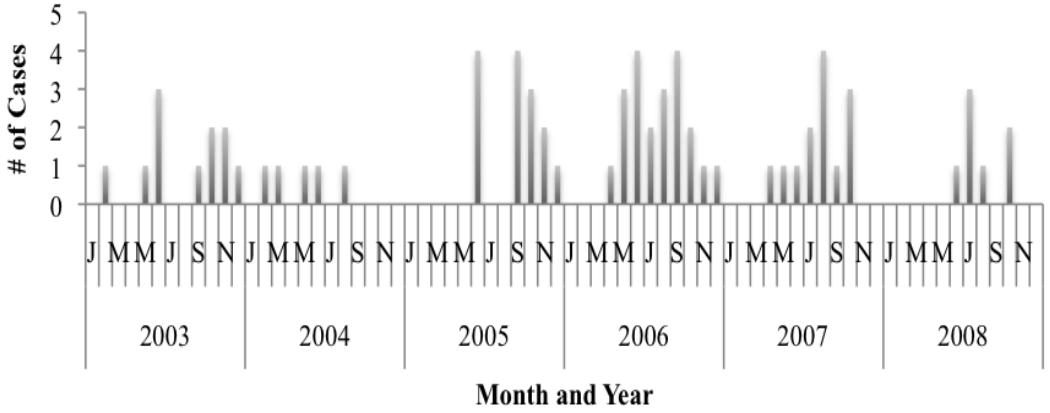


Figure 8f: Mpumalanga

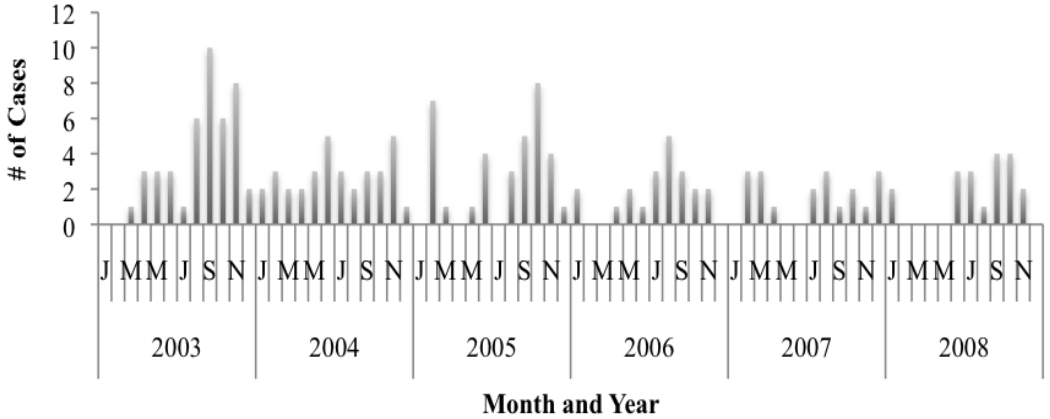


Figure 8g: Northern Cape

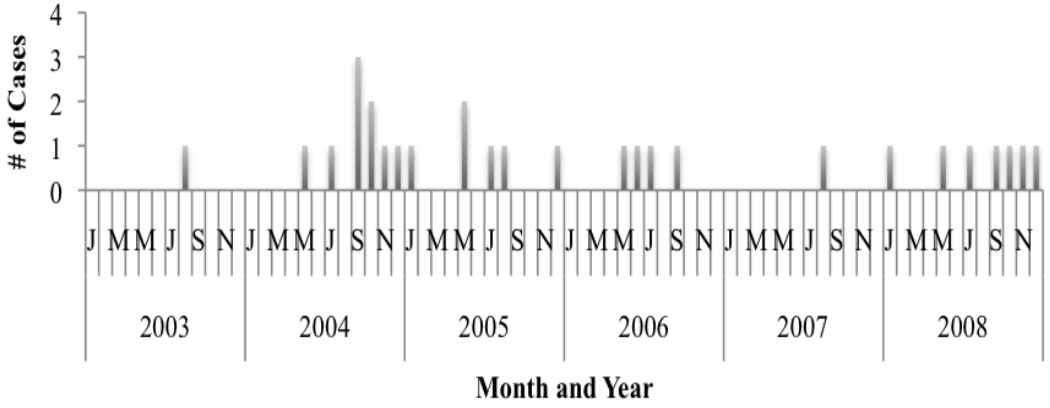


Figure 8h: North West

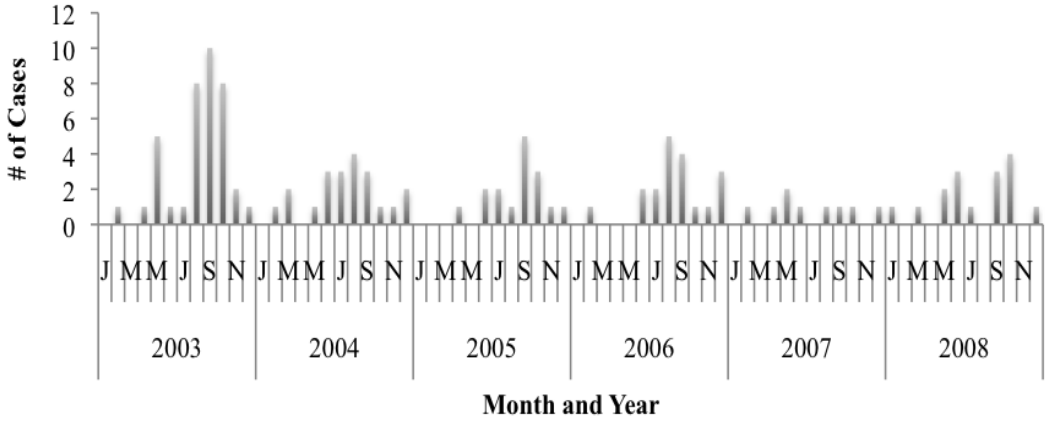


Figure 8i: Western Cape

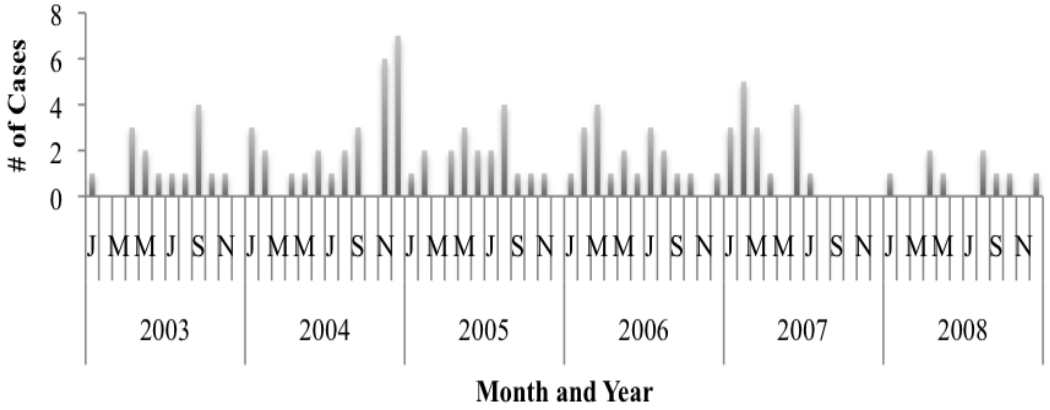


Table 1

Association between cases of Serotype 1 IPD and prior HIV infection, as well as other Covariates

Variable		Number of IPD cases (%)	Serotype 1 IPD (%)	Control Pool* (%)	OR (95% CI)	p-value
HIV status	Infected	3003 (80)	432 (71)	2571 (82)	0.54 (0.45, 0.67)	<0.01
	Uninfected	753 (20)	177 (29)	576 (18)		
	Total	3756	609	3327		
Age group	≤2	2877 (31)	90 (5)	2787 (38)	referent	<0.01
	3 to 4	993 (11)	120 (6)	873 (12)	4.26 (3.21, 5.66)	
	5 to 17	1048 (11)	418 (22)	630 (9)	20.55 (16.10, 26.21)	
	18 to 24	400 (4)	173 (9)	227 (3)	23.60 (17.69, 31.49)	
	25 to 45	2757 (30)	822 (43)	1935 (27)	13.15 (10.50, 16.48)	
	46 to 64	879 (10)	242 (13)	637 (9)	11.76 (9.10, 15.21)	
	≥65	211 (2)	47 (2)	164 (2)	8.87 (6.03 13.06)	
	Total	9165	1912	7253		
Antibiotic Use in Past 2 Months	Yes	626 (17)	26 (4)	600 (20)	0.17 (0.11, 0.25)	<0.01
	No	2962 (83)	607 (96)	2355 (80)		
	Total	3588	633	2955		
Nosocomial Infection	Yes	229 (2)	26 (1)	203 (2)	0.43 (0.29, 0.65)	<0.01
	No	11668 (98)	2662 (99)	9006 (98)		
	Total	11897	2688	9209		
Underlying Condition	none	9449 (95)	2014 (97)	7435 (94)	referent	<0.01
	protein malnutrition	288 (3)	14 (1)	274 (3)	0.19 (0.11, 0.32)	
	cardiac condition	97 (1)	23 (1)	74 (1)	1.14 (0.72, 1.84)	
	pulmonary condition	113 (1)	12 (1)	101 (2)	0.44 (0.24, 0.80)	
	Total	9947	2063	7884		
Lifestyle	neither	1638 (86)	288 (80)	1350 (87)	referent	<0.01
	smoker	54 (3)	13 (3)	41 (3)	1.49 (0.79, 2.81)	
	alcohol user	139 (7)	35 (10)	104 (7)	1.58 (1.05, 2.36)	
	both	82 (4)	28 (7)	54 (3)	2.43 (1.51, 3.90)	
	Total	1913	364	1549		
Race	Asian	21 (0.2)	8 (0.4)	13 (0.2)	2.20 (0.91, 5.33)	<0.01
	Black	9715 (95)	2120 (96.1)	7595 (95.8)	referent	
	Coloured	346 (3.4)	33 (1.5)	313 (4.0)	0.38 (0.26, 0.54)	
	White	146 (1.4)	43 (2.0)	103 (1.0)	1.59 (1.04, 2.14)	
	Total	10228	2204	8024		
Province	EC	590 (5)	101 (4)	489 (5)	0.56 (0.45, 0.70)	<0.01
	FS	706 (6)	181 (7)	525 (6)	0.93 (0.78, 1.11)	
	GA	6211 (51)	1679 (62)	4532 (47)	referent	
	KZ	1542 (13)	242 (9)	1300 (14)	0.50 (0.43, 0.58)	
	LP	248 (2)	71 (3)	177 (2)	1.08 (0.82, 1.43)	
	MP	589 (5)	170 (6)	419 (4)	1.10 (0.91, 1.32)	
	NC	121 (1)	29 (1)	92 (1)	0.85 (0.56, 1.30)	
	NW	379 (3)	121 (4)	258 (3)	1.30 (1.01, 1.58)	
	WC	1665 (13)	108 (4)	1557 (17)	0.19 (0.15, 0.23)	
	Total	12051	2702	9349		
Gender	Female	5947 (50)	1281 (49)	4666 (51)	0.90 (0.83, 0.99)	0.02
	Male	5842 (50)	1360 (51)	4483 (49)		
	Total	11789	2641	9149		

*14, 6A, 6B, 23F, 19A, 19F

Table 2

ARV Therapy and Cotrimoxazole Prophylaxis Among HIV Positive Patients with Cases of IPD

Variable	Frequency (%)	Serotype 1 IPD	Control Pool* (%)	OR (95% CI)	p-value
ARV					
Current Therapy	318 (15)	8 (3)	310 (17)	0.14 (0.07, 0.28)	<0.01
No Therapy	1752 (85)	279 (97)	1473 (83)		
Total	2070	287	1783		
Cotrimoxazole Prophylaxis	940 (44)	51	889	0.24 (0.18, 0.33)	<0.01
No Prophylaxis	1208 (56)	230	978		
Total	2148	281	1867		

*14, 6A, 6B, 23F, 19A, 19F

Table 3

Association of Prior Serotype 1 IPD with Disease Outcome, Severity, and Clinical Syndrome

Variable		Frequency (%)	Serotype 1 IPD (%)	Control Pool* (%)	OR (95% CI)	p-value
Disease outcome	mortality	1580 (29)	318 (31)	1262 (28)	1.14 (0.99, 1.32)	0.08
	survival	3901 (71)	705 (69)	3196 (72)		
	Total	5481	1023	4458		
Disease severity	severe	201 (5)	46 (6)	155 (5)	1.42 (1.01, 2.00)	0.04
	mild/moderate	3843 (95)	663 (94)	3180 (95)		
	Total	4044	709	3335		
Meningitis	Yes	2659 (43)	688 (52)	1971 (41)	1.56 (1.38, 1.76)	<0.01
	bacteremia or LRTI	3523 (57)	645 (48)	2878 (60)		
	Total	6182	1333	4849		
LRTI	Yes	3213 (52)	606 (45)	2607 (54)	0.72 (0.63, 0.80)	<0.01
	bacteremia or meningitis	2969 (48)	727 (55)	2242 (46)		
	Total	6182	1333	4849		
Bacteremia w/o focus	Yes	310 (5)	39 (3)	271 (6)	0.51 (0.36, 0.72)	<0.01
	LRTI or meningitis.	5872 (95)	1294 (97)	4578 (94)		
	Total	6182	1333	4849		

*14, 6A, 6B, 23F, 19A, 19F

Table 4

Multivariable analysis of risk factors for an outcome of serotype 1 IPD relative to control pool*, conditioned on race

Variable		Odds ratio	95% CI	p-value
HIV status	infected	0.23	0.15, 0.36	<0.01
	uninfected			
Age group	≤2	referent		
	3 to 4	10.69	3.90, 29.29	<0.01
	5 to 17	78.73	31.61, 196.08	<0.01
	18 to 24	64.05	23.72, 172.95	<0.01
	25 to 45	39.51	16.21, 96.27	<0.01
	46 to 64	21.54	8.01, 57.94	<0.01
	≥65	28.12	5.65, 139.98	<0.01
Lifestyle	neither	referent		
	smoker	1.34	0.61, 2.93	0.46
	alcohol user	1.50	0.87, 2.58	0.14
	both	2.02	1.05, 3.90	0.04

*14, 6A, 6B, 23F, 19A, 19F

Table 5

Multivariable analysis of association between antiretroviral therapy and an outcome of serotype 1 IPD among HIV-infected patients, relative to control pool*, conditioned on year

Variable		Odds ratio	95% CI	p-value
ARV Therapy	current	0.23	0.07, 0.80	0.02
	none			
Cotri Prophylaxis	Yes	0.32	0.17, 0.60	<0.01
	No			
Lifestyle	neither	referent		
	smoker	1.51	0.58, 3.91	0.40
	alcohol user	1.70	0.84, 3.44	0.14
	both	2.20	1.00, 4.86	0.05
Age Category	≤2	referent		
	3 to 4	17.09	1.93, 150.99	0.02
	5 to 17	99.76	13.08, 760.96	<0.01
	18 to 24	51.49	6.49, 408.27	<0.01
	25 to 45	38.81	5.27, 285.66	<0.01
	46 to 64	12.44	1.43, 108.06	0.02
	≥65	44.70	3.14, 636.91	<0.01

*14, 6A, 6B, 23F, 19A, 19F

Table 6

Multivariable analysis of serotype 1 IPD as a predictor of mortality relative to survival, conditioning on age

Variable		Odds ratio	95% CI	p-value
IPD	serotype 1	0.82	0.60, 1.12	0.22
	control pool*			
HIV status	infected	2.38	1.76, 3.21	<0.01
	uninfected			
Underlying condition	none	referent		
	protein malnutrition	1.45	0.99, 2.14	0.06
	pulmonary condition	23.51	9.27, 59.65	<0.01
	cardiac condition	0.79	0.39, 1.60	0.51
Meningitis	Yes	3.24	2.62, 4.02	<0.01
	LRTI or bacteremia			

*14, 6A, 6B, 23F, 19A, 19F

Table 7

Multivariable analysis of serotype 1 IPD as a predictor of severe IPD, relative to mild/moderate

Variable		Odds ratio	95% CI	p-value
IPD	serotype 1	1.44	0.97, 2.15	0.07
	control pool*			
Nosocomial Infection	Yes	2.19	1.18, 4.08	0.01
	No			
LRTI	Yes	0.41	0.30, 0.58	<0.01
	Meningitis or bacteremia			
Age Category	≤2	referent		
	3 to 4	0.61	0.27, 1.38	0.23
	5 to 17	1.03	0.55, 1.93	0.94
	18 to 24	1.63	0.73, 3.65	0.24
	25 to 45	1.52	0.97, 2.38	0.07
	46 to 64	1.90	1.09, 3.31	0.02
	≥65	4.17	1.94, 8.96	<0.01

*14, 6A, 6B, 23F, 19A, 19F

Table 8

Multivariable analysis of serotype 1 IPD as a predictor of meningitis relative to LRTI and bacteremia without focus, conditioned on province and race

Variable		Odds ratio	95% CI	p-value
IPD	serotype 1	0.60	0.43, 0.85	<0.01
	control pool*			
HIV status	infected	0.56	0.44, 0.72	<0.01
	uninfected			
Underlying Condition	none	referent		
	protein malnutrition	0.52	0.35, 0.80	<0.01
	cardiac condition	1.57	0.67, 3.70	0.30
	pulmonary condition	0.28	0.11, 0.71	<0.01
Antibiotic use in past 2 months	Yes	0.91	0.67, 1.23	0.53
	No			

*14, 6A, 6B, 23F, 19A, 19F

Table 9

Multivariable analysis of serotype 1 IPD as a predictor of LRTI relative to meningitis and bacteremia without focus, conditioned on race and province

Variable		Odds ratio	95% CI	p-value
IPD	serotype 1	1.42	1.01, 2.00	0.04
	control pool*			
HIV status	infected	1.91	1.49, 2.45	<0.01
	uninfected			
Underlying Condition	none	referent		
	protein malnutrition	1.51	1.04, 2.21	<0.01
	cardiac condition	0.27	0.11, 0.66	0.04
	pulmonary condition	2.64	1.27, 5.47	<0.01
Age Category	≤2	referent		
	3 to 4	1.66	1.18, 2.33	<0.01
	5 to 17	0.95	0.67, 1.36	0.79
	18 to 24	1.40	0.77, 2.54	0.27
	25 to 45	1.63	1.20, 2.21	<0.01
	46 to 64	1.53	0.96, 2.43	0.07
	≥65	9.30	1.06, 81.75	0.04
Antibiotic use in past 2 months	Susceptible	1.30	0.97, 1.74	0.07
	Non-susceptible			

*14, 6A, 6B, 23F, 19A, 19F

Table 10

Multivariable analysis of serotype 1 IPD as a predictor of bacteremia without focus relative to LRTI and meningitis

Variable		Odds ratio	95% CI	p-value
IPD	serotype 1	1.01	0.60, 1.71	0.96
	control pool*			
HIV	infected	0.40	0.28, 0.56	<0.01
	uninfected			
Underlying Condition	none	referent		
	protein malnutrition	1.84	1.11, 3.05	0.02
	cardiac condition	2.07	0.85, 5.09	0.11
	pulmonary condition	1.00	0.43, 2.33	1.00

*14, 6A, 6B, 23F, 19A, 19F